PATTERN RECOGNITION IN THE DEVELOPING MAIZE LEAF EPIDERMIS: GENE NETWORK ANALYSES AND MACHINE LEARNING APPROACHES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Pengfei Qiao

August 2019

PATTERN RECOGNITION IN THE DEVELOPING MAIZE LEAF EPIDERMIS: GENE NETWORK ANALYSES AND MACHINE LEARNING APPROACHES

Pengfei Qiao, Ph. D.

Cornell University 2019

ABSTRACT

Biological systems contain data of high dimensions and magnitudes, including biochemistry, cellular patterning, transcriptomics, and genomics. Here I combined network analyses and machine learning, to identify developmental patterns that may be amenable to the improvement of drought tolerance. Cuticles comprise the hydrophobic lipid layer covering the aboveground plant body, and have long been a research focus into water conservation in plants. However, no prior studies have examined cuticle development across a temporal and spatial gradient in a crop plant. I used gene network analyses to correlate the biochemical/developmental gradient of cuticle components with the underlying transcriptomic transitions to identify the role of PHYTOCHROME B-mediated light signaling in cuticle development. Subsequent statistical and biochemical analysis revealed LIPID-TRANSFER PROTEINs as evolutionary novelties contributing to the emergence of cuticles in land plants. Additionally, combining the power of genome- and transcriptome-wide association studies (GWAS and TWAS), vesicular trafficking was implicated in the regulation of

cuticular evaporation rate. Water loss through the leaf surface is also moderated by specialized cell types (bulliform cells) in maize. Bulliform cell ontogeny was investigated in the developing maize leaf, and a machine learning approach (convolutional neural networks) was employed to conduct high-throughput phenotyping of microscopic bulliform cell traits in 60,780 leaf epidermal glue-impression images. A subsequent GWAS analysis on bulliform cell column number and column width identified a set of gene candidates implicated to function in cell division and DNA methylation. Overall this dissertation demonstrates a multidisciplinary approach combining developmental biology, transcriptomics, quantitative genetics, machine learning, and statistical data analysis, toward a more holistic understanding of the mechanisms of water conservation in maize.

BIOGRAPHICAL SKETCH

Pengfei Qiao came to Cornell in the year of 2015. He later joined Dr. Michael Scanlon's lab, where he spent all his Ph.D. career using data analytics investigating water conservation in plants. Pengfei focuses on the interface between a plant and its environment – the epidermis, and conducted his research from two perspectives: cuticles and bulliform cells. He developed great interest in using data analytics and machine learning to answer biological questions, and did so successfully. Pengfei cares deeply about how plants respond to abiotic stresses such as drought, and hopes his work will inspire future action on water crisis.

ACKNOWLEDGMENTS

I owe many thanks to all the people who have helped and inspired me throughout this journey, and I am afraid I cannot list them all here. I would like to begin by saying thank you all, I would not be here if it were not for you.

I would like to thank my advisor, Mike Scanlon. I am forever in your debt for the kind and intelligent mentoring you have given me. Up to this day I still remember the first time I walked into your office trying to rotate in your lab, and I never left. You have taught me science and how to do research, and you did an excellent job in that. On top of being a good academic advisor, you have also been an awesome life mentor. I learned communication, negotiation, kindness, patience, not only in being a scientist with research skills, but more importantly, on good scientist ethics. You have also been most supportive on the important decisions I have made, and truly want the best for my development. Borrowing the words from a graduate of the lab – you are a true student advocate. You said that once you become a graduate student's advisor, you are their advisor for life. Your tremendous help and support have helped me thus far, and I would love to continue to have such an awesome advisor for the many years to come. I want to say, very sincerely, from the bottom of my heart, thank you, Mike. My appreciation is like Cayuga's waters, deep and wide.

I would like to thank my committee members, Mike Gore, Isabel Molina, and Mert Sabuncu. Mike, all the discussions I have had with you were such enjoyments. I admire your diligence and great personality, and appreciate all the help and time you have devoted to my research. You are a great mentor, and a great friend. Isabel, you have been so kind and supportive of all the hypotheses I made in my research. You

devoted so much time in our collaboration, and you have provided benevolent help. Every time I doubt my ideas, I think of your role model of the brave scientist that you have set. Mert, although we did not have many in-person discussions, you have given me some of the most interesting and creative ideas in my graduate career. Every time we meet it was like I was reading an entirely new book – exciting, inspiring, and challenging. I appreciate you agreeing to provide so much help and support for my ideas, and I hope to be an amazing scientist someday, just like you.

Almost no research can be done alone, and without my collaborators' devotion, this dissertation would have been much shallower. Laurie, I have to admit I was intimidated by you when we first met in January 2016! I later realized that you are not only a serious and inspiring scientist, but also a supportive mentor. Thank you for all the heated discussions, and for your support in conducting all the experiments described in this dissertation. I would also like to thank you for leading such an amazing and intelligent team of PIs on this giant NSF grant, and for organizing all the project meetings and logistics. Susanne, Miguel, Albert, you have been my precious friends and scientific fellows, and I really appreciate your contributions on my research. Isabel and Richard, thanks for all your help in analyzing the numerous cuticle samples I have sent. Running GC-MS is not trivial, neither is the time commitment in doing so. You have both been so supportive, and have never said no to even the silliest request. I would have discovered nothing, if it were not for you. Mike and Meng, you are like my co-advisors. You were never tired of me bothering you with the most minute GWAS and statistics questions. Meng, you have also listened to all my complaints when I felt the most anxious. I might seem to have taken it all for granted, but I truly am grateful.

I would also like to thank Cornell University for all their resources. I was able to just walk across the street to drop off sequencing samples, and never had to go outside in the winter to visit my precious plants in the Weill growth chamber. The awesome networking opportunities you host, the rigorous and diverse courses you provide, have really helped me thus far. You sincerely care about your students' well-being, and "hail, all hail, Cornell!".

Hosted inside Cornell is my beloved graduate program. Cornell *does* have the best Plant Biology program, where each of our individual needs and ambitions are respected and nurtured. I found the best group of faculty in this program, a supportive administrative staff, and an amazing cohort of fellow graduate students. Intelligent minds gather here, and our conversations sparkle with joy and wisdom.

My friends, including my cohort, have been by my side. You have made Ithaca a better place, without you, it would never have been this fun. You brought laughter, joy, and excitement. From the golf course to the garden plot, from de Tasty to Saigon Kitchen, and from California to Vermont, I spent some of the best years in my life with you guys, and I will never regret it.

My lab has some of my most important friends, and here I would like to thank you all for the discussions of research, the help and guidance, and the tolerance of my sometimes annoying behavior. I enjoy spending time in the lab, and you all have been an important part of my life.

Finally, I would like to thank NSF for their generous research funding. You enabled the discovery of new biological phenomena, as well as helped my personal development.

Thank you all!

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH V

ACKNOWLEDGMENTS VI

CHAPTER 1: INTRODUCTION 1

1.1 OVERVIEW 1
1.2 CUTICLES 2
1.2.1 THE ROLE OF CUTICLES IN PLANT EVOLUTION AND ADAPTATION 2
1.2.2 CUTICLE STRUCTURE AND COMPOSITION 3
1.2.3 CUTICLE BIOSYNTHESIS AND ITS REGULATION 4
1.2.4 CUTICULAR EVAPORATION/TRANSPIRATION 7
1.3 BULLIFORM CELLS 8
1.4 QUANTITATIVE GENETICS 12
1.4.1 GENOME-WIDE ASSOCIATION STUDY AND LINKAGE DISEQUILIBRIUM 12
1.4.2 HERITABILITY 13
1.4.3 LINEAR MODELS 14
1.4.4 TRANSCRIPTOME-WIDE ASSOCIATION STUDY 15
1.5 GENE NETWORK ANALYSIS 16
1.5.1 OVERVIEW 16
1.5.2 GENE CO-EXPRESSION NETWORK 18
1.5.3 GENE REGULATORY NETWORKS 20
1.6 CONVOLUTIONAL NEURAL NETWORKS 20
1.6.1 FORWARD PROPAGATION AND BACKPROPAGATION 21
1.6.2 NETWORK ARCHITECTURE OF A U-NET 24
1.7 THESIS OUTLINE: CHAPTER PUBLICATIONS AND AUTHOR CONTRIBUTIONS 26
1.8 REFERENCES 28

CHAPTER 2: NETWORK ANALYSES IMPLICATE A ROLE FOR PHYTOCHROMES AND LIPID TRANSFER PROTEINS IN THE EVOLUTION OF PLANT CUTICLES 41

2.1 ABSTRACT 41
2.2 SIGNIFICANCE STATEMENT 42
2.3 INTRODUCTION 42
2.4 RESULTS AND DISCUSSION 44
2.4.1 TRANSCRIPTOMIC ANALYSES OF CUTICLE DEVELOPMENT IN THE ADULT MAIZE LEAF 45
2.4.2 IDENTIFICATION OF PUTATIVE CUTICLE REGULATORY GENES ACROSS THE LEAF DEVELOPMENTAL GRADIENT VIA DIRECTED NETWORK INFERENCE 46
2.4.3 A WEIGHTED CO-EXPRESSION NETWORK ANALYSIS IDENTIFIES ADDITIONAL CANDIDATE REGULATORY GENES FOR CUTICLE BIOSYNTHESIS 53
2.4.4 LIGHT-REGULATED CUTICLE DEVELOPMENT: *PHYTOCHROME* (*PHY*) MUTANTS HAVE ALTERED CUTICLE COMPOSITION 58
2.5 MATERIALS AND METHODS 80
2.5.1 PLANT MATERIAL AND GROWTH CONDITIONS 80
2.5.2 LASER MICRODISSECTION AND RNASEQ ANALYSIS 81

2.5.3 WAX EXTRACTION AND ANALYSIS 81
2.5.4 DIFFERENTIAL EXPRESSION ANALYSIS 82
2.5.5 WEIGHTED CO-EXPRESSION NETWORK ANALYSIS 82
2.5.6 CASUAL STRUCTURE INFERENCE ANALYSIS 82
2.5.7 TWO-WAY ANOVA 83
2.5.8 TRANSGENIC LINES 83
2.6 ACKNOWLEDGMENTS 83
2.7 REFERENCES 85

CHAPTER 3: GENOMIC ANALYSES IDENTIFY A ROLE FOR VESICULAR TRAFFICKING DURING CUTICULAR EVAPORATION IN MAIZE LEAVES 93

3.1 ABSTRACT 93
3.2 INTRODUCTION 93
3.3 MATERIALS AND METHODS 95
3.3.1 EXPERIMENTAL DESIGN 95
3.3.2 TISSUE COLLECTION AND RNA SEQUENCING 96
3.3.3 PHENOTYPIC DATA COLLECTION 97
3.3.4 STATISTICAL ANALYSIS 98
3.3.5 GENOTYPIC DATA IMPUTATION 99
3.3.6 GENOME-WIDE ASSOCIATION STUDY 100
3.3.7 TRANSCRIPTOME-WIDE ASSOCIATION STUDY 100
3.3.9 FISHER'S COMBINED TESTS 101
3.4 RESULTS AND DISCUSSION 101
3.4.1 PHENOTYPIC VARIABILITY OF CUTICULAR EVAPORATION RATE 101
3.4.2 GENOME-WIDE ASSOCIATION STUDY IDENTIFIES MVB TRAFFICKING GENES AS POTENTIAL CE RATE REGULATORS 103
3.4.3 TRANSCRIPTOME-WIDE ASSOCIATION STUDY SUGGESTS THAT THE CE RATE IS TRANSCRIPTIONALLY REGULATED 107
3.4.4 FISHER'S COMBINED PROBABILITY TEST SELECTS CANDIDATES IMPLICATED IN CUTICLE BIOSYNTHESIS AND MVB TRAFFICKING 109
3.4.5 CAMOCO-AIDED GENE DISCOVERY 112
3.5 REFERENCES 114

CHAPTER 4: MACHINE LEARNING ENABLES HIGH-THROUGHPUT PHENOTYPING FOR ANALYSES OF THE GENETIC ARCHITECTURE OF BULLIFORM CELL PATTERNING IN MAIZE 122

4.1 ABSTRACT 122
4.2 INTRODUCTION 122
4.3 MATERIALS AND METHODS 125
4.3.1 BULLIFORM CELL ONTOGENY AND RNA SEQUENCING 125
4.3.2 DIFFERENTIAL GENE EXPRESSION ANALYSIS 126
4.3.3 EXPERIMENTAL DESIGN 126
4.3.4 LEAF EPIDERMAL PHENOTYPIC DATA COLLECTION 127
4.3.5 NEURAL NETWORKS IN THE QUANTIFICATION OF PHENOTYPIC DATA 128
4.3.6 STATISTICAL DATA ANALYSIS 129
4.3.7 DNA EXTRACTION, GENOTYPING AND SNP IDENTIFICATION 130
4.3.8 GENOME-WIDE ASSOCIATION STUDY 131
4.3.9 LINKAGE DISEQUILIBRIUM ANALYSIS 132

4.3.10 DATA AVAILABILITY 132
4.4 RESULTS AND DISCUSSION 132
4.4.1 BULLIFORM CELL ONTOGENY 132
4.4.2 PHENOTYPE VARIABILITY AND PHENOTYPING ACCURACY 135
4.4.3 GWAS OF BULLIFORM CELL PATTERNING TRAITS 139
4.5 ACKNOWLEDGEMENTS 145
4.6 REFERENCES 146

CHAPTER 5: CONCLUSION 154

5.1 REFERENCES 158

CHAPTER 6 APPENDIX: TRANSCRIPTOMIC ANALYSES OF GENES REGULATING SAM SIZE IN MAIZE 161

6.1 ABSTRACT 161
6.2 INTRODUCTION 161
6.3 MATERIALS AND METHODS 162
6.3.1 PLANT MATERIAL AND GROWTH CONDITIONS 162
6.3.2 LASER-MICRODISSECTION RNA SEQUENCING (LM-RNASEQ) 162
6.3.3 DIFFERENTIAL GENE EXPRESSION ANALYSIS 162
6.4 RESULTS AND DISCUSSION 163
6.4.1 DIFFERENTIAL GENE EXPRESSION ANALYSIS 163
6.5 REFERENCES 168

BIBLIOGRAPHY 170

Chapter 1: Introduction

1.1 Overview

Land plants (embryophytes) were presented with significant challenges when they colonized he terrestrial environment more than 450 million years ago, including increases in photoradiation, supporting the weight of the plant body in a non-aqueous environment, and drastic decreases in water potential encountered on the arid Earth. The epidermis comprises the outermost layer of the plant body, and is the interface between plants and the external environment. Thus, the epidermal surface of land plants performs a critical function in protection from water loss (Kenrick and Crane, 1997a; Bateman *et al.*, 1998b; Raven and Edwards, 2004; Kerstiens, 2006b; Jetter and Riederer, 2016a).

The cuticle is a pivotal adaptation in embryophyte evolution. Cuticles comprise a hydrophobic complex of lipid polymers and waxes that forms a protective film sealing the majority of the shoot epidermis, preventing water evaporation from non-stomatal plant surfaces (Jetter and Riederer, 2016a). In addition to the cuticle, some plants have evolved specific epidermal cell types that function to cope with water loss. In the grasses for example, bulliform cells are enlarged parenchymatous cells on the adaxial (top) epidermis that enable leaf rolling during heat stress and drought. Rolling reduces the surface area of leaves exposed to the sun, and hence reduces water loss (Hsiao *et al.*, 1984; Price *et al.*, 1997; Becraft *et al.*, 2002b; Dai *et al.*, 2007; Kadioglu and Terzi, 2007; Bennetzen and Hake, 2008; Hu *et al.*, 2010; Robinson *et al.*, 2015).

Numerous studies have identified many of the biosynthetic and transport genes involved in plant cuticle biogenesis. Reverse genetics has also identified a small number of genes that function in bulliform cell patterning (Dai *et al.*, 2007; Fujino *et al.*, 2008; Itoh *et al.*, 2008b; Hibara *et al.*, 2009b; Zhang *et al.*, 2009a; Li *et al.*, 2010b; Zou *et al.*, 2011; Fang *et al.*, 2012b; Xiang *et al.*, 2012; Chen *et al.*, 2015a). However, such studies are lacking on agronomically important U. S. crops. In this thesis, we used the maize adult leaf as a model system to study cuticle development, water loss from cuticles (cuticular transpiration), and bulliform cell patterning. Combining high-dimensional transcriptomic, phenotypic, biochemical, quantitative genetic, and genomic data with machine learning and statistical analyses, this cross-disciplinary study aims to identify biological patterns in the developing maize leaf contributing to water conservation and drought tolerance. This study generated a variety of unique, cell/tissue specific transcriptomic datasets that are available to the maize genetics community. Specifically, we provide novel insight into the evolutionary-developmental aspects of cuticle development, and the genetic architecture of bulliform cell pattering and non-stomatal water loss via cuticular evaporation in maize leaves.

1.2 Cuticles

1.2.1 The role of cuticles in plant evolution and adaptation

Cuticles evolved as a means to reduce water loss in the embryophytes; green algal ancestors of the land plants live in aquatic environments. Because they are submerged in water, green algae do not face desiccation challenges (Kenrick and Crane, 1997a; Bateman *et al.*, 1998b; Raven and Edwards, 2004). As plants moved from aqueous

environments to the arid land, water loss became a predominant issue for the embryophytes. Land plants evolved multiple strategies to cope with this new challenge. Some plants developed enlarged cells that function as water storage units (Mauseth, 1995; Ogburn and Edwards, 2010), whereas others developed deep and wide root systems (Ashton, 1975; Monneveux and Belhassen, 1996; Coppen, 2003) and still others developed morphological adaptations to decrease the surface area exposed to the desiccating air (Boke, 1980). However, among the various water conservation mechanisms found within the embryophytes, the hydrophobic, extra-epidermal cuticle is a conserved evolutionary innovation that is found in *all* land plants, but not in the green algae (Kenrick and Crane, 1997a; Bateman *et al.*, 1998b; Raven and Edwards, 2004).

1.2.2 Cuticle structure and composition

The embryophyte body is composed of three tissue types: dermal tissue, ground tissue, and vascular tissue (Edwards, 1993; Raven *et al.*, 2005). Dermal tissue, specifically the epidermis, is the outermost layer covering the plant body. As the primary interface between a plant and its surroundings, the epidermis serves important protective functions (Edwards, 1993; Raven *et al.*, 2005). To augment this epidermal protection and prevent excessive water loss, an extracellular, hydrophobic lipid layer known as the cuticle covers the entire aboveground shoot. Due to its hydrophobic property, the cuticle protects inner cells from water loss (Raven and Edwards, 2004; Kerstiens, 2006b; Yeats and Rose, 2013a; Jetter and Riederer, 2016a).

Cuticles are composed of lipids deposited on epidermal cell walls and are interlocked with cell wall polysaccharides to form a matrix of lipid polymers. This dense polymer

is called cutin; cutins are polymerized fatty acids connected by primary and secondary ester bonds (Kolattukudy, 2001a; Nawrath, 2002b; Yeats and Rose, 2013a). Because all biochemical analyses of cutin composition involve a depolymerization step, such that the precise macromolecular structure of cutin is not fully understood (Yeats and Rose, 2013a; Fich *et al.*, 2016).

Intracuticular waxes are molecules embedded within the cutin matrix. Unlike cutins, waxes consist of long-chain, non-polar carbon molecules that are soluble in organic solvents such as chloroform. Typical wax components include hydrocarbons, aldehydes, alcohols, ketones, and wax esters. Epicuticular waxes are composed of similar carbon molecules, but are not embedded in the cutin matrix and can be removed with adhesives such as gum Arabic. Epicuticular waxes can form epicuticular wax crystals and films, which often render a glaucous or glossy appearance to the cuticle surface (Kolattukudy, 2001a; Nawrath, 2002b; Yeats and Rose, 2013a).

Cuticle composition can differ drastically among different plant species. For example, in *Arabidopsis thaliana* leaf waxes contain only trace amounts of wax esters (Broun *et al.*, 2004; Li *et al.*, 2008), whereas in adult maize leaf cuticles wax esters are predominant cuticle components (Bourgault *et al.*, 2019b). Few studies have examined different stages of cuticle ontogeny within the same species. Thus, questions remain: Is there a cuticle gradient during ontogeny and what are the mechanism contributing to this gradient? This dissertation addresses these questions in Chapter 2.

1.2.3 Cuticle biosynthesis and its regulation

Cuticles consist of cutins and waxes, and are synthesized within plant epidermal cells (Suh *et al.*, 2005a; Kunst and Samuels, 2009b). In *Arabidopsis thaliana*, long-chain

C_{16} and C_{18} fatty acids are synthesized in epidermal cell plastids and subsequently transported into the endoplasmic reticulum (ER), and finally converted into CoA thioesters by LONG-CHAIN ACYL-COENZYME A SYNTHASE (LACS). In the ER, the fatty acid elongase (FAE) complex converts these thioesters into very-long-chain fatty acid CoAs (Suh *et al.*, 2005a; Joubès *et al.*, 2008; Kunst and Samuels, 2009b), whereupon they enter one of two pathways. In the reduction pathway, the fatty acid CoAs are reduced by Fatty Acyl-CoA Reductase (FARs) to form free, primary alcohols, which could also be esterified with fatty acids to form wax esters. Alternatively, the long-chain fatty acids may enter the decarbonylation pathway leading to the formation of aldehydes, alkanes, secondary alcohols, or ketones. The biosynthesis of the cutin monomer 10,16-dihydroxyhexadecanoic acid also begins with fatty acids derived from epidermal plastids. Following the formation of C16-CoA, successive ω-hydroxylation (by members of the cytochrome subfamily CYP86, such as CYP86A4 in Arabidopsis), mid-chain hydroxylation (by the cytochrome subfamily CYP77, such as CYP77A6 in Arabidopsis) and acyl-transfer (by LACS), acyl-glycerols are generated (Yeats and Rose, 2013a). Both waxes and cutin monomers are exported in parts into the epidermal, outer cell wall via ATP-BINDING CASSETTE G (ABCG) transporters located on the plasma membrane (Yeats and Rose, 2013a), although other transport mechanisms have also been described. Within the apoplast, cutin monomers are polymerized into a matrix by cutin synthases such as GDSL1, and waxes are further transported to the outermost layer of the cuticle by LIPID-TRANSFER PROTEINs (LTPs) (Debono *et al.*, 2009; Lee *et al.*, 2009; Kim *et al.*, 2012). The biosynthesis and transportation of cuticles can be summarized in Figure 1.1.

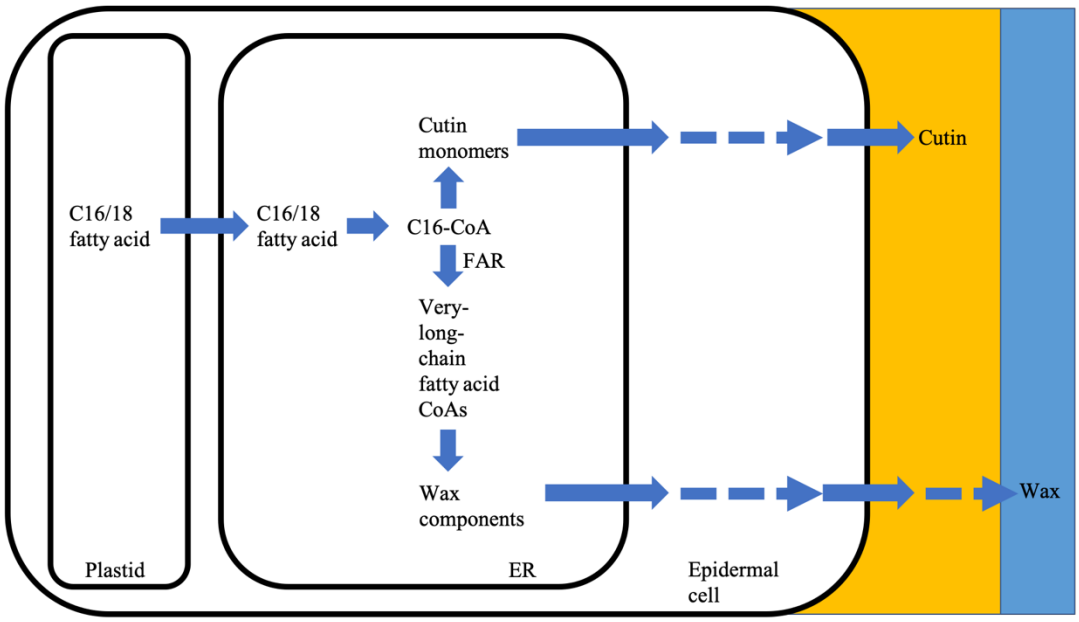

Figure 1.1. An illustration of cuticle biogenesis. The orange box represents cutin matrix that is interlocked with the cell wall, and the blue box represents the wax components deposited on top of the cutin matrix. For simplicity, waxes embedded in the cutin matrix is not shown, but the transition from cutin matrix to waxes are gradual. Dashed arrows indicate processes that are not fully understood, and solid arrows indicate conversion of chemical substances or their transportation.

The regulation of cuticle biosynthesis is complex. In *Arabidopsis thaliana*, abscisic acid (ABA), water deficit, and osmotic stress all induce the formation of cuticular wax. In response to environmental stress, ABA activates the expression of MYB96 transcription factor, which induce wax biosynthesis. Likewise, the overexpression of members of the WAX INDUCED1/SHINE1 (WIN1/SHN1) family of transcription factors leads to increased production of wax and cutin, as well as the induction of several pectin-modifying enzymes. These concerted activities of WIN1/SHIN1 suggest that the biosynthesis of cuticles and cell walls may be co-regulated, such that the cuticle is regarded as a specialized extension of the plant cell wall. Other transcription factors regulating cuticle biosynthesis include MYB106 and MYB16, which directly activate *WIN1/SHN1*, and MYB30, which responds to pathogen attacks. (Hooker *et al.*, 2002; Suh *et al.*, 2005a; Yeats and Rose, 2013a; Suh and Go, 2014)

Beyond transcription factors, post-transcriptional regulation is also involved in cuticle biosynthesis. For example, *AtCER7* encodes an exosomal exoribonuclease, and has been proposed to be involved in the degradation of small RNAs that regulate *CER3* transcript levels (Hooker *et al.*, 2007; Lam *et al.*, 2012). Additional examples of post-translational regulation examples include CER9. In Arabidopsis, *CER9* encodes a protein involved in the protein degradation process, and has been proposed to maintain the homeostasis of cuticle biosynthesis enzymes (Lü *et al.*, 2012).

1.2.4 Cuticular evaporation/transpiration

Water transpiration through the leaf is referred to as “conductance.” In an ideal environment with optimal humidity, the stomatal conductance is denoted as g_{max}, wherein the “max” subscript indicates that transpiration is at its maximum. When

drought stressed or in the dark, most plants have the ability to close their stomata. Upon stomata closure water loss becomes minimal, and is denoted as g_{min} (Kelliher *et al.*, 1995; Daszkowska-Golec and Szarejko, 2013). In our study, we use the term "cuticular evaporation," or CE, to represent the water loss through the cuticle. Note that we do not assume that all stomata are completely sealed.

Water movement from inside the leaf to the surrounding air is entirely passive, and driven by differences in water potential. Water potential measure the abundance of water molecules in a particular environment or system. Water potential difference (gradient) drives water movement toward spaces with lower water potential, in most cases, from the inside of the leaf to the surface (Cowan, 1965; Molz, 1981; Boyer, 1985; Raven *et al.*, 2005). Since the loss of water is entirely passive, water will also escape through the leaf epidermis in the absence of some preventative barrier. The hydrophobic properties of cuticle molecules minimize this water movement, but cuticles are not a complete water seal. When water transpiration through the stomata is minimized, such as when stomata are closed during a hot, summer day, water loss through the cuticle is maximized and the plant is vulnerable to drought stress (Raven *et al.*, 2005). Hence, the cuticle comprises the primary barrier to water loss from the maize shoot, and the optimization of CE rate is desirable for agricultural production.

1.3 Bulliform cells

Grasses can roll their leaves inward, toward the adaxial surface (top), when exposed to severe drought and heat stress. The leaf rolling response reduces air flow and photoradiation on the leaf surface, thereby physically reducing water potential differences between the epidermis and the inner leaf layers. This mitigation in water

potential gradient directly reduces water loss and plant desiccation (Becraft *et al.*, 2002b; Bennetzen and Hake, 2008).

It has been long hypothesized that leaf rolling is dependent upon a specialized, adaxial epidermal cell type – the bulliform cell. Bulliform cells are morphologically distinct from all other epidermal cell types, owing to their large volume, rounded cell shape, and the presence of unique cuticular ridges on their surface. Bulliform cells appear in parallel rows along the proximodistal axis of the adaxial leaf blade. It is postulated that bulliform cells lose water and shrink in size, mechanically folding the adaxial leaf surface inward to effect leaf rolling. In grasses, bulliform cell ontogeny follows a linear pattern along the proximodistal axis. Therefore, bulliform cells align in an almost parallel and linear fashion from the basal (proximal) end of the leaf to the tip (distal) region (Figure 1.2). This enables the leaf to roll perpendicularly to the proximodistal axis, reducing the exposed surface area (Hsiao *et al.*, 1984; Price *et al.*, 1997; Becraft *et al.*, 2002b; Dai *et al.*, 2007; Kadioglu and Terzi, 2007; Bennetzen and Hake, 2008; Hu *et al.*, 2010; Robinson *et al.*, 2015).

Recent research on bulliform cell development has been focused in two, strategic areas in rice: (1) transcriptomics of bulliform cells and (2) reverse genetics and mutant screening. In one transcriptomic study, bulliform cells express ~ 16,000 genes, more than the median of 8,831 in the 40 cell/tissue types sampled (Jiao *et al.*, 2009). This indicates that the developmental patterning of bulliform cells may be polygenic. Multiple mutants also have been identified with altered bulliform cell phenotypes. For example, mutations in the *RICE OUTERMOST CELL-SPECIFIC GENE5 (Roc5)* of rice condition increases in the number of bulliform cells per row, as well as bulliform

cell area. The increased number of bulliform cells also conditions a more upright leaf angle in *roc5* mutants, which is a desirable agronomic trait (Zou *et al.*, 2011). Fujino *et al.* showed that *narrow leaf 7* mutants contain more bulliform cells per column, but these cells are significantly smaller in size than in wild type siblings (Fujino *et al.*, 2008). Yet another study found that *rolled and erect leaf* 1 mutant leaves have more abundant *and* larger bulliform cells (Chen *et al.*, 2015a). In addition, mutations in *leaf inclination2*, *abaxially curled leaf 1*, and *semi-rolled leaf1* also result in increased numbers of bulliform cells in rice (Li *et al.*, 2010b; Zhao *et al.*, 2010; Xiang *et al.*, 2012). Moreover, some mutations affecting dorsiventral leaf polarity also disrupt

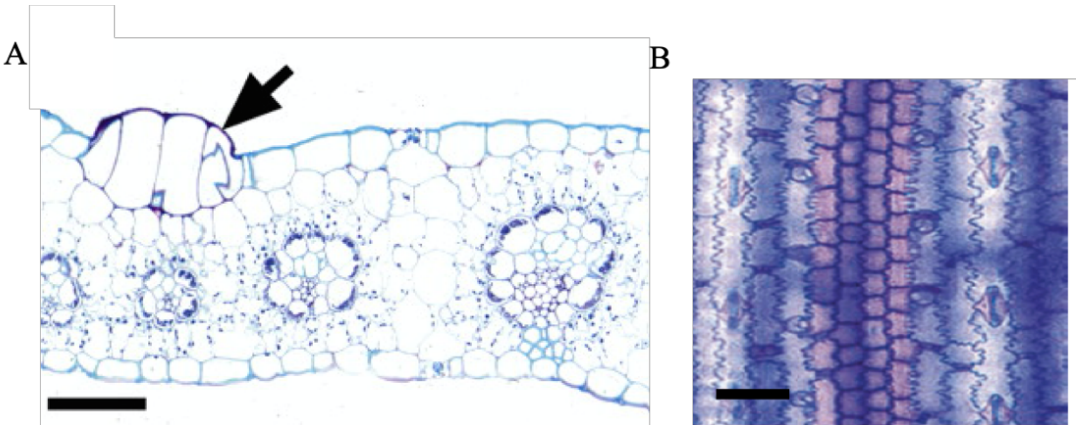

Figure 1.2. Images of bulliform cells shown in a cross section (A) and epidermal peel (B). (A) black arrow head indicates a cluster of bulliform cells in the adaxial surface. (B) An epidermal peel of adult maize leaf shows a linear column of bulliform cells. Scale bar 100 μm. Figures adapted from (Becraft *et al.*, 2002b; Bennetzen and Hake, 2008).

bulliform cell patterning. Hu *et al.* (2010) observed that the leaves of *narrow and rolled leaf1* mutants have reduced bulliform cell size. Finally, the leaf blades of the rice mutants s*hallot-like1*, *adaxialized leaf1*, and *rolled leaf1* develop ectopic bulliform-like cells on the abaxial surface (Hibara *et al.*, 2009b; Zhang *et al.*, 2009b). As yet, there are no investigations of the genes contributing to the phenotypic diversity in bulliform cell patterning in maize. In Chapter 4 of this dissertation, we use quantitative genetic approaches to explore the genetic architecture underlying this microscopic phenotype.

1.4 Quantitative genetics

1.4.1 Genome-wide association study and linkage disequilibrium

Understanding the genetic basis of phenotypic variation has been of long-standing interests to biologists. In recent years, advances in modern quantitative genetics and technology have enabled the identification of genes contributing to quantitative traits in high-throughput. Genome Wide Association Study, or GWAS, is one example wherein high-throughput sequencing is utilized to generate high-density genetic markers for use in quantitative genetic analysis. GWAS uses historical recombination within a population to find associations between genetic markers and prospective trait-of-interest. In this way, potential casual genes can be identified as contributing to the phenotypic variation of a specific trait (Nordborg and Weigel, 2008; Korte and Farlow, 2013).

Recombination occurs during meiosis, after homologous chromosomes align to form the synaptonemal complex. Subsequently, cross-overs may occur between

homologous chromosomes, enabling the exchange of genetic materials to form recombinant chromosomes. In this way, recombination is the foundation for GWAS. Traditionally, when two markers are on the same chromosome, and are *not separated* by genetic recombination at a frequency of more than 50%, they are said to be "linked." "Linkage disequilibrium (LD)", similarly, is defined as the non-random association of alleles at different loci in a population. LD blocks (chromosomal regions in LD) are generally shorter in outbreeding species such as maize relative to selfers, enabling a high mapping resolution (Remington *et al.*, 2001; Visscher *et al.*, 2012; Visscher *et al.*, 2017).

In GWAS, single nucleotide polymorphisms (SNPs) are widely used as genetic markers. SNPs can be identified at relatively low cost, and these markers are stable across generations. Thus, the goal for GWAS is to find significant associations between SNPs and a phenotype of interest (Nordborg and Weigel, 2008; Korte and Farlow, 2013).

1.4.2 Heritability

Each observed trait is the sum of three discrete components: genetics, environment, and the interaction between these two factors. If we use P to denote the phenotype, G as the genotypic component, and E as the environmental effect, we can write this phenotype decomposition as:

$$P = G + E + G \times E$$

The identification of potential causal SNPs contributing to the trait of interest in a GWAS population requires that the trait has consistency across different trials, i.e.

heritable. The (broad-sense) heritability is defined as the fraction of the variance in the phenotype that results from G. A high heritability would indicate a strong and consistent underlying genetic architecture regulating the trait, which is desirable for the mapping of causal SNPs using GWAS (Lynch and Walsh, 1998; Bernardo, 2002; Holland *et al.*, 2003; Hung *et al.*, 2012).

1.4.3 Linear models

Early in the 20th century, the genetic distance was calculated by hand for each individual genetic marker, due to the small number of markers. In the modern genomics era, megabytes of marker (SNP) data are available, and it is not possible to calculate manually the association of each SNP to a given trait. With such a high magnitude of genotypic data, statistical caveats also became non-trivial. The ideal GWAS model has a minimum false positive rate (Type I error rate) and maximum statistical power to detect real associations (1 – Type II error rate). There is also the problem of multiple testing on every SNP in the population, where the overall Type I error rate needs to be controlled rather than each individual test (Johnson *et al.*, 2010).

The most naïve linear model used for GWAS analysis is $Y \sim X\beta + \epsilon$, where Y is the phenotype onto which the SNPs (X) are regressed. Each SNP is tested individually and independently; SNPs with significant correlation to the phenotype (significant β) are selected, and candidate genes are selected if they are in LD with these significant SNPs.

One common problem with this naïve model is that it gives an inflated Type I error rate. Type I error rate inflation could be caused by population structure associated with the trait. Maize GWAS populations usually consists of different subpopulations that

cluster together based on similar genetic structure, likely due to germplasm resources, shared ancestry, regional adaptation and breeding histories. Different subpopulations may have distinct phenotypic distributions for the trait-of-interest. For example, tropical maize lines generally flower later than temperate lines. The shift of trait distributions among subpopulations can lead to false positive identification of significant alleles, whereas the diverse alleles associated with different subpopulations are falsely identified as potential causal alleles of a particular trait. Hence the inclusion of population structure as covariates in the GWAS model is usually desired, or at least should be tested within the model (Yu *et al.*, 2006; Zhang *et al.*, 2010; Visscher *et al.*, 2012; Visscher *et al.*, 2017).

While the use of population structure covariates in the GWAS model can alleviate the problem of Type I error rate inflation, this might be insufficient or may over-correct the model. To address this issue, Yu *et al.* (2006) included the relatedness between individual lines (kinship) in their GWAS model, such that the phenotypic variance that is truly caused by SNPs can be dissected out. Models containing both population structure and kinship are often referred to as the popular Q+K method. Such a unified mixed linear model has improved control of Type I error rate, and the power to identify genetic signals contributing to the trait (Yu *et al.*, 2006; Zhang *et al.*, 2010).

1.4.4 Transcriptome-wide association study

GWAS is powerful at detecting SNP in strong LD with causal variants for the trait-of-interest, but in practice, the presence of multiple genes within an LD block can complicate the nomination of casual gene candidates. Similar to GWAS, where high magnitudes of SNP data are generated, massively parallel sequencing can rapidly

generate Big transcriptomic Data. The various mRNAs expressed in a given cell/tissue/organ are indicative of biological function. Thus, correlating the trait-of-interest with the associated transcriptome can identify potential causal genes (van 't Veer *et al.*, 2002; Hirsch *et al.*, 2014; Gyorffy *et al.*, 2015; Lin *et al.*, 2017; Pasaniuc and Price, 2017; Kremling *et al.*, 2018a).

As in GWAS, a TWAS can benefit from linear models, where the gene expression levels are regressed onto the trait. However, one caveat is that the gene expression level *alone* is a phenotype. As described in section 1.4.2, the environment may exert a drastic influence on gene expression levels, which in turn biased estimates of the gene's effect. In addition to the environment, human error may contribute to batch variation during RNA extraction and sequencing library preparation, whereas lane effect may generate more variation during sequencing. These factors may introduce hidden structures within the data that are difficult to model. A Probabilistic Estimate of Expression Residuals (PEER) has been proposed to remove the "noise" in the data. The PEER framework takes raw transcriptomic data and models these "hidden factors", outputting residuals that are more reflective of the "true" transcriptome (Stegle *et al.*, 2010; Stegle *et al.*, 2012; Kremling *et al.*, 2018a). TWAS and GWAS results can be aggregated and analyzed together using a Fisher's combined test, which elevates the power to detect potential causal genes for the trait of interest (Fisher, 1925; Li *et al.*, 2014).

1.5 Gene network analysis

1.5.1 Overview

The concept of a network is not limited to biological systems. Perhaps more intuitively, airports are connected to each other by flights traveling in between cities, and people are connected to each other in professional or social networks (Barabasi, 2003; Barabasi, 2013). Similarly, biological molecules regulate each other by forming biological networks (Barabasi and Oltvai, 2004). The most central nodes comprise the "hubs" of the network, such as a busy airport, or a popular influencer on social media. There are multiple ways to measure "centrality", or how central a node is in a given network. For example, the number of connections is an indicator of the centrality of a node, as is the calculation of how frequently a certain node must be traversed in order to connect any other two nodes. Thus, hubs are vital to the function of the entire network; removal of hubs may have detrimental effects. In a biological system, the hubs can function as "master regulators" of a certain process. For example, one transcription factor may activate and/or repress the expression of multiple genes with diverse functions within the network. As a result, elimination of a single transcription factor functioning as the hub of a network may inactivate the entire network (Barabasi, 2003; Barabasi and Oltvai, 2004; Barabasi, 2013).

In addition to the nodes, other important constituents of a network are the edges, which connect the nodes together. Edges may represent different concepts in different networks. For example, person A and B may be connected by friendship within a social network, thereby creating an edge. Likewise, the flow of information from person C to person D can be represented as an edge. Note that in these examples, the edges represent distinct concepts. Friendship between two people has no direction, which is the signature of an "undirected network". On the other hand, the example of information flow described above is directional, flowing from person C to person D

but not vice versa. Hence information flow is a “directed network”. Examples of undirected and directed network are demonstrated in Figure 1.3.

In the following two sections, the concepts of undirected and directed networks will be applied to biological systems, namely plant transcriptomes (Barabasi, 2003; Barabasi, 2013).

1.5.2 Gene co-expression network

Given the large scale of transcriptomic data generated by RNA Sequencing (RNAseq), there are multiple ways to model these data and extract meaningful biological insight. One method is to construct a gene co-expression network. The basic idea behind gene co-expression networks is that genes involved in the same pathway, perhaps gene regulatory or biosynthetic, tend to display similar expression patterns across biological samples. As suggested by the name, the edges in this undirected gene co-expression network area represent the degree at which any two genes are “co-expressed” in the same biological space and time (Langfelder and Horvath, 2008; Langfelder *et al.*, 2008; Horvath, 2011; Langfelder and Horvath, 2012). After the construction of a co-expression network, it becomes feasible to identify the hub genes regulating the biological processes of interest.

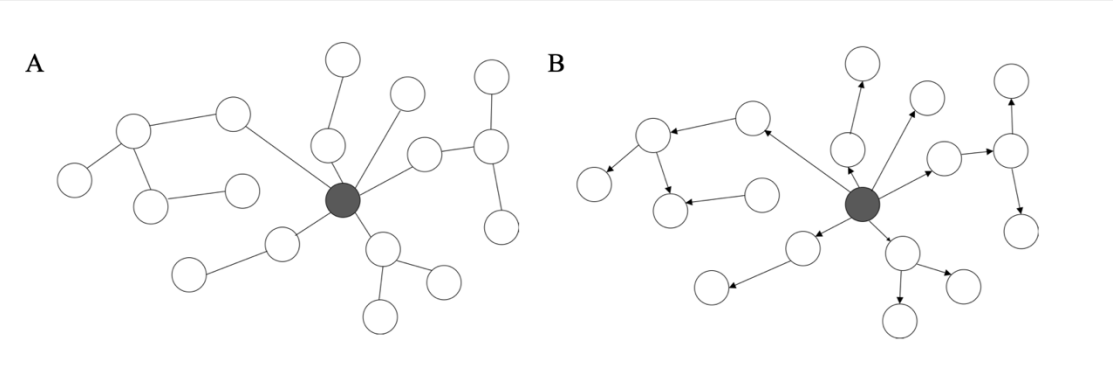

Figure 1.3. Demonstrations of an undirected network (A) and directed network (B). Each circle represents a node, and the lines connecting the circles present edges. The black node is the hub in both networks.

1.5.3 Gene regulatory networks

Unlike gene co-expression networks, a gene regulatory network (GRN) is directed. The edges of a GRN point from a "parent" gene, such as a transcription factor, to its target, such as an enzyme-encoding gene for fatty acid elongation. The edge can be modeled in different ways, for example, Penfold *et al.* proposed a method to calculate the probability of observing one target gene's expression c*onditional on* the expression level of its transcriptional regulator (Penfold and Wild, 2011). This powerful approach considers the direction of gene-to-gene interactions, although it is computationally intensive and thus only feasible for a small set of genes.

1.6 Convolutional neural networks

An artificial intelligence (AI) system recognizes patterns, and extrapolates these patterns without human interference (Wilson, 2014). Raw input can be represented as tensors (multi-dimensional geometric vectors), which are used to train and test the AI system to recognize patterns. Once trained, the AI system can take intelligent actions and make accurate predictions without human intervention.

An early machine learning model, called perceptron, was inspired by human neural systems. A perceptron from 1962 calculated a linear combination of the input values (inner product), which was updated every time the perceptron made a mistake on the training sample (Block *et al.*, 1962). Later in 1998, LeCun *et al.* built a convolutional neural network (CNN) to recognize hand-written digits (LeCun *et al.*, 1998). However, challenges such as the lack of training data and constraints on computational hardware restricted its application (Deng *et al.*, 2009; Russakovsky *et al.*, 2015).

The size of image data has exploded within recent years, which requires an efficient and accurate algorithm for visual data interpretation. The early 2000's were marked by the arrival of benchmark datasets such as PASCAL and ImageNet, which enabled vast increases in the sizes of AI training sets and provided an important building block for the computer vision field (Deng *et al.*, 2009; Everingham *et al.*, 2010; Russakovsky *et al.*, 2015). In the year 2012, the CNN algorithm AlexNet won the ImageNet challenge with a dramatically improved error rate (Krizhevsky *et al.*, 2012). On the other hand, advances in computational hardware such as graphics processing units (GPUs) enabled the computation of CNNs with millions of parameters in a matter of minutes.

At present, CNNs are used for object classification, semantic segmentation, activity recognition and more (LeCun *et al.*, 1995; LeCun *et al.*, 1998; Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014; Zeiler and Fergus, 2014; Ronneberger *et al.*, 2015; Szegedy *et al.*, 2015; He *et al.*, 2016). In Chapter 4 of this thesis, a particular CNN model called the U-net is used for segmenting leaf-impression light micrographs to enable high-throughput phenotyping of a microscopic trait, owing to its fast computation and efficient use of context pixel information.

1.6.1 Forward propagation and backpropagation

In essence, CNNs are very similar to perceptrons. Starting from the input X, information is passed forward into the hidden layers of the neural network. Figure 1.4 shows an example of a fully connected network with three hidden layers. After

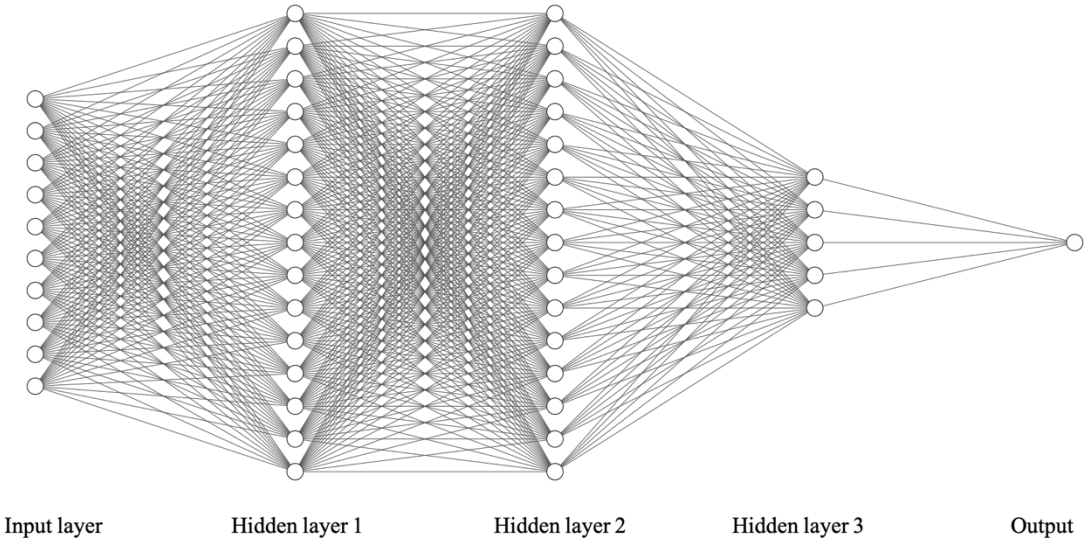

Figure 1.4. An illustration of a fully-connected neural network with three hidden layers. Each circle represents a node, and all the nodes in a certain layer are connected with all the nodes in the previous layer. Each connection represents an edge, where a linear function of the incoming node value is propagated into the target node. A non-linear function is applied to the input value of the target node subsequently. After the third hidden layer, a weighted sum of the nodes in hidden layer 3 is summarized as the output.

calculating the inner product between the input of each layer and the corresponding coefficients (weights) of that layer (a linear function), the result is passed onto an activation function (Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014).

One major advantage of neural networks is the inclusion of non-linearity in their activation functions. Non-linear layers enable the modeling of non-linear patterns in the data, and are more versatile in real life examples. Examples of popular choices of activation functions are sigmoid function, tanh, and rectified linear units (ReLUs) (Goodfellow *et al.*, 2016).

Common layers include convolutional layers, max pooling layers, and fully connected layers. In a convolutional layer, weights are shared across all segments of the input tensor, and multiple weights are trained. Once trained, a max pooling step selects the maximum values of each segment. After multiple convolutional and max pooling layers, usually in the last layer, all nodes are connected with every node from the previous layer, hence the name "fully connected" (Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014; Ronneberger *et al.*, 2015).

After the information flows through the entire network, the difference between the output and the true value, or the cost function, can provide useful gradient information to optimize the network. The gradient is the derivative of the functions multiplied by some scalar (learning rate). Using the chain rule to calculate the derivative of the composite function, gradients can be calculated backward from the output all the way back to the first hidden layer, hence the name "backpropagation". The gradient is used to optimize the network via algorithms such as stochastic gradient descent (Hecht-Nielsen, 1992). It is worth mentioning that there is no guarantee that the objective

function in a neural network is convex; as such there is no guarantee that a global optimum can be reached through learning. However, in practice the local optima could be comparable to the global optima in terms of its performance(Blum and Rivest, 1989; Kawaguchi, 2016).

1.6.2 Network architecture of a U-net

A U-net is a type of neural net that consists of a contracting stage, followed by an expanding phase, hence the name U-net. In the contracting phase, the raw image is provided as the input, going through convolutional layers and max pooling layers, with increased number of filters and decreased size of the output. Later in the expanding stage, several up-convolutional layers are combined with convolutional layers partially copied from the contracting stage, to increase the size of the image back to its original dimensions. The output is a segmentation map of the raw input image. The contracting-expanding path enables the algorithm to utilize the context information of a certain pixel in the input image, which facilitates the classification of that pixel and decreases the required size of the training set (Ronneberger *et al.*, 2015). The architecture of u-net is shown in Figure 1.5.

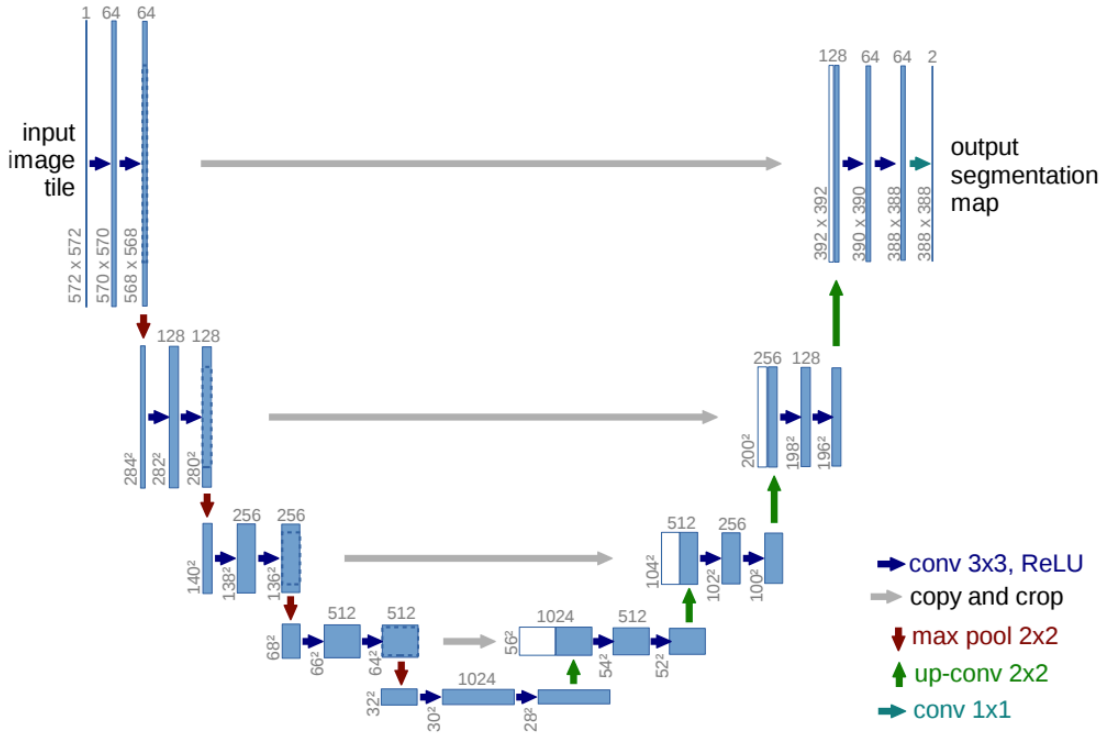

Figure 1.5. U-net architecture. Each box represents a layer, and horizontal numbers on top of the boxes are the number of corresponding layers. Different arrows indicate different operations between layers. Vertical numbers represent the shape of the arrays in corresponding layers. Figure adapted from (Ronneberger *et al.*, 2015).

1.7 Thesis outline: Chapter publications and author contributions

Chapter 2 of this dissertation is devoted to deciphering the potential regulators of cuticle biosynthesis in an expanding adult maize leaf. All the biochemical data described in this chapter was generated by Dr. Richard Bourgault, Marc Mohammadi and Dr. Isabel Molina. I generated all transcriptomic data and performed the downstream data analysis. This manuscript is currently under revision for resubmission at *Proceeding of the National Academy of Science, USA* (*PNAS*).

The cuticle plays an important role in preventing water loss, however, plants still lose water through CE. The third chapter of this dissertation focuses on the identification of CE rate regulators. A TWAS was performed on CE rates that were measured in the Wisconsin Diversity panel by collaborators in the Smith lab and the Gore lab. Leaf tissue was also collected by the Smith lab the Gore lab. The TWAS data was combined with GWAS and co-expression analyses to provide more supporting material for the casual models. Dr. Meng Lin in the Gore lab and I are co-first authors on this manuscript. The samples were prepared by Dr. Lin and myself. I processed the RNSseq reads and performed the co-expression analysis, and validated the TWAS and GWAS results from Dr. Lin. A manuscript is in preparation for submission to *The Plant Cell*.

Chapter 4 focuses on dissecting the genetic architecture of bulliform cell patterning. I collected tissues, made and took images of leaf impressions, and performed high-throughput phenotyping of this microscopic trait with CNNs, mentored by Dr. Mert Sabuncu. Dr. Michael Scanlon performed analyses of bulliform cell ontogeny. We observed that bulliform cell traits such as row number and row width are highly

heritable, and GWAS was performed on these traits, with the help of Dr. Meng Lin and Dr. Mike Gore. Dr. Lin calculated the BLUEs and heritabilities, and I performed the GWAS. This manuscript is to be submitted for publication in *Genetics*.

Appendix 1 describes a laser-microdissection RNAseq analysis of genes regulating shoot apical meristem (SAM) size variation in maize. This work was performed during my first semester rotation project in the Scanlon lab, and will be published as part of a larger paper describing a GWAS analysis of SAM morphology performed by former graduate student Dr. Sam Leiboff. I will appear as a co-author on this manuscript.

Overall, this dissertation provides a unique angle to study the maize leaf epidermis. Since embryophytes first invaded the land more than 450 million years ago, water loss has been a persistent challenge. Fluctuations in precipitation, and extremes in ambient temperature combined with the parching sun, all pose the threat of dehydration and desiccation. In this rapidly-changing world, the degree of environmental instability requires plants to be more adept at dealing with water loss. This evolutionarily "old" challenge is still present, yet ever-changing. I hope my study will contribute to the improvement of agricultural systems, to engineer crops to adapt to drought stress in the present, and in the future.

1.8 REFERENCES

Ashton, D.J.A.J.o.B. (1975). The root and shoot development of Eucalyptus regnans F. Muell **23,** 867-887.

Barabasi, A.-L. (2003). Linked: How everything is connected to everything else and what it means.

Barabasi, A.L. (2013). Network science. Philos Trans A Math Phys Eng Sci **371,** 20120375.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet **5,** 101-113.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., Stein, W.E.J.A.R.o.E., and Systematics. (1998). Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation **29,** 263-292.

Becraft, P.W., Li, K., Dey, N., and Asuncion-Crabb, Y.J.D. (2002). The maize dek1 gene functions in embryonic pattern formation and cell fate specification **129,** 5217-5225.

Bennetzen, J.L., and Hake, S.C. (2008). Handbook of maize: its biology. (Springer Science & Business Media).

Bernardo, R. (2002). Breeding for quantitative traits in plants. (Stemma press Woodbury).

Block, H., Knight Jr, B., and Rosenblatt, F.J.R.o.M.P. (1962). Analysis of a four-layer series-coupled perceptron. II **34,** 135.

Blum, A., and Rivest, R.L. (1989). Training a 3-node neural network is NP-complete. In Advances in neural information processing systems, pp. 494-501.

Boke, N.H.J.B. (1980). Developmental morphology and anatomy in Cactaceae **30,** 605-610.

Bourgault, R., Matschi, S., Vasquez, M., Qiao, P., Sonntag, A., Charlebois, C., Mohammadi, M., Scanlon, M.J., Smith, L.G., and Molina, I.J.b. (2019). Changes in lipid composition and ultrastructure associated with functional maturation of the cuticle during adult maize leaf development**,** 625343.

Boyer, J.S.J.A.r.o.p.p. (1985). Water transport **36,** 473-516.

Broun, P., Poindexter, P., Osborne, E., Jiang, C.-Z., and Riechmann, J.L.J.P.o.t.N.A.o.S. (2004). WIN1, a transcriptional activator of epidermal wax accumulation in Arabidopsis **101,** 4706-4711.

Chen, Q., Xie, Q., Gao, J., Wang, W., Sun, B., Liu, B., Zhu, H., Peng, H., Zhao, H., and Liu, C.J.J.o.e.b. (2015). Characterization of Rolled and Erect Leaf 1 in regulating leave morphology in rice **66,** 6047-6058.

Coppen, J.J. (2003). Eucalyptus: the genus Eucalyptus. (CRC Press).

Cowan, I.J.J.o.A.E. (1965). Transport of water in the soil-plant-atmosphere system**,** 221-239.

Dai, M., Zhao, Y., Ma, Q., Hu, Y., Hedden, P., Zhang, Q., and Zhou, D.-X.J.P.p. (2007). The rice YABBY1 gene is involved in the feedback regulation of gibberellin metabolism **144,** 121-133.

Daszkowska-Golec, A., and Szarejko, I. (2013). Open or close the gate - stomata action under the control of phytohormones in drought stress conditions. Front Plant Sci **4,** 138.

Debono, A., Yeats, T.H., Rose, J.K., Bird, D., Jetter, R., Kunst, L., and Samuels, L. (2009). Arabidopsis LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. Plant Cell **21,** 1230-1238.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (Ieee), pp. 248-255.

Edwards, D.J.N.P. (1993). Cells and tissues in the vegetative sporophytes of early land plants **125,** 225-247.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., and Zisserman, A.J.I.j.o.c.v. (2010). The pascal visual object classes (voc) challenge **88,** 303-338.

Fang, L., Zhao, F., Cong, Y., Sang, X., Du, Q., Wang, D., Li, Y., Ling, Y., Yang, Z., and He, G.J.P.b.j. (2012). Rolling-leaf14 is a 2OG-Fe (II) oxygenase family protein that modulates rice leaf rolling by affecting secondary cell wall formation in leaves **10,** 524-532.

Fich, E.A., Segerson, N.A., and Rose, J.K.J.A.R.o.P.B. (2016). The plant polyester cutin: biosynthesis, structure, and biological roles **67,** 207-233.

Fisher, R.A.J.E., Scotland. (1925). Statistical methods for research workers Oliver and Boyd **6**.

Fujino, K., Matsuda, Y., Ozawa, K., Nishimura, T., Koshiba, T., Fraaije, M.W., Sekiguchi, H.J.M.G., and Genomics. (2008). NARROW LEAF 7 controls leaf shape mediated by auxin in rice **279,** 499-507.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. (MIT press).

Gyorffy, B., Hatzis, C., Sanft, T., Hofstatter, E., Aktas, B., and Pusztai, L. (2015). Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Res **17,** 11.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In Neural networks for perception (Elsevier), pp. 65-93.

Hibara, K.-i., Obara, M., Hayashida, E., Abe, M., Ishimaru, T., Satoh, H., Itoh, J.-i., and Nagato, Y.J.D.b. (2009). The ADAXIALIZED LEAF1 gene functions in leaf and embryonic pattern formation in rice **334,** 345-354.

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2014). Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26,** 121-135.

Holland, J.B., Nyquist, W.E., and Cervantes-Martínez, C.T.J.P.b.r. (2003). Estimating and interpreting heritability for plant breeding: an update **22,** 9-112.

Hooker, T.S., Millar, A.A., and Kunst, L. (2002). Significance of the expression of the CER6 condensing enzyme for cuticular wax production in Arabidopsis. Plant Physiol **129,** 1568-1580.

Horvath, S. (2011). Weighted network analysis: applications in genomics and systems biology. (Springer Science & Business Media).

Hsiao, T.C., O'Toole, J.C., Yambao, E.B., and Turner, N.C.J.P.P. (1984). Influence of osmotic adjustment on leaf rolling and tissue death in rice (Oryza sativa L.) **75,** 338-341.

Hu, J., Zhu, L., Zeng, D., Gao, Z., Guo, L., Fang, Y., Zhang, G., Dong, G., Yan, M., and Liu, J.J.P.m.b. (2010). Identification and characterization of NARROW ANDROLLED LEAF 1, a novel gene regulating leaf morphology and plant architecture in rice **73,** 283-292.

Hung, H., Browne, C., Guill, K., Coles, N., Eller, M., Garcia, A., Lepak, N., Melia-Hancock, S., Oropeza-Rosas, M., and Salvo, S.J.H. (2012). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population **108,** 490.

Itoh, J.-I., Hibara, K.-I., Sato, Y., and Nagato, Y.J.P.P. (2008). Developmental role and auxin responsiveness of Class III homeodomain leucine zipper gene family members in rice **147,** 1960-1975.

Jetter, R., and Riederer, M.J.P.p. (2016). Localization of the transpiration barrier in the epi-and intracuticular waxes of eight plant species: water transport resistances are associated with fatty acyl rather than alicyclic components **170,** 921-934.

Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H.Y., Zhao, H., Deng, X.W., and Nelson, T. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet **41,** 258-263.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A., and O'Brien, S.J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics **11,** 724.

Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R.J.P.m.b. (2008). The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling **67,** 547.

Kadioglu, A., and Terzi, R.J.T.B.R. (2007). A dehydration avoidance mechanism: leaf rolling **73,** 290-302.

Kawaguchi, K. (2016). Deep learning without poor local minima. In Advances in neural information processing systems, pp. 586-594.

Kelliher, F., Leuning, R., Raupach, M., Schulze, E.-D.J.A., and Meteorology, F. (1995). Maximum conductances for evaporation from global vegetation types **73,** 1-16.

Kenrick, P., and Crane, P.R.J.N. (1997). The origin and early evolution of plants on land **389,** 33.

Kerstiens, G.J.J.o.E.B. (2006). Water transport in plant cuticles: an update **57,** 2493-2499.

Kim, H., Lee, S.B., Kim, H.J., Min, M.K., Hwang, I., and Suh, M.C. (2012). Characterization of glycosylphosphatidylinositol-anchored lipid transfer protein 2 (LTPG2) and overlapping function between LTPG/LTPG1 and LTPG2 in cuticular wax export or accumulation in Arabidopsis thaliana. Plant Cell Physiol **53,** 1391-1403.

Kolattukudy, P.E. (2001). Polyesters in higher plants. In Biopolyesters (Springer), pp. 1-49.

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. Plant Methods **9,** 29.

Kremling, K., Diepenbrock, C., Gore, M., Buckler, E., and Bandillo, N.J.b. (2018). Transcriptome-wide association supplements genome-wide association in Zea mays**,** 363242.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097-1105.

Kunst, L., and Samuels, L.J.C.o.i.p.b. (2009). Plant cuticles shine: advances in wax biosynthesis and export **12,** 721-727.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9,** 559.

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw **46**.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics **24,** 719-720.

LeCun, Y., Bengio, Y.J.T.h.o.b.t., and networks, n. (1995). Convolutional networks for images, speech, and time series **3361,** 1995.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.J.P.o.t.I. (1998). Gradient-based learning applied to document recognition **86,** 2278-2324.

Lee, S.B., Go, Y.S., Bae, H.J., Park, J.H., Cho, S.H., Cho, H.J., Lee, D.S., Park, O.K., Hwang, I., and Suh, M.C. (2009). Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules, and enhanced susceptibility to infection by the fungal pathogen Alternaria brassicicola. Plant Physiol **150,** 42-54.

Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., and Kunst, L.J.P.p. (2008). Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis **148,** 97-107.

Li, L., Shi, Z.-Y., Li, L., Shen, G.-Z., Wang, X.-Q., An, L.-S., and Zhang, J.-L.J.M.p. (2010). Overexpression of ACL1 (abaxially curled leaf 1) increased bulliform cells and induced abaxial curling of leaf blades in rice **3,** 807-817.

Li, Q., Hu, J., Ding, J., and Zheng, G. (2014). Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. Biostatistics **15,** 284-295.

Lin, H.Y., Liu, Q., Li, X., Yang, J., Liu, S., Huang, Y., Scanlon, M.J., Nettleton, D., and Schnable, P.S. (2017). Substantial contribution of genetic variation in the expression of transcription factors to phenotypic variation revealed by eRD-GWAS. Genome Biol **18,** 192.

Lynch, M., and Walsh, B. (1998). Genetics and analysis of quantitative traits. (Sinauer Sunderland, MA).

Mauseth, J.D.J.B.o.t.T.B.C. (1995). Collapsible water-storage cells in cacti**,** 145-151.

Molz, F.J.J.W.r.r. (1981). Models of water transport in the soil-plant system: A review **17,** 1245-1260.

Monneveux, P., and Belhassen, E. (1996). The diversity of drought adaptation in the wide. In Drought Tolerance in Higher Plants: Genetical, Physiological and Molecular Biological Analysis (Springer), pp. 7-14.

Nawrath, C.J.T.A.b.A.S.o.P.B. (2002). The biopolymers cutin and suberin **1**.

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. Nature **456,** 720-723.

Ogburn, R.M., and Edwards, E.J. (2010). The ecological water-use strategies of succulent plants. In Advances in botanical research (Elsevier), pp. 179-225.

Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet **18,** 117-127.

Penfold, C.A., and Wild, D.L. (2011). How to infer gene networks from expression profiles, revisited. Interface Focus **1,** 857-870.

Price, A.H., Young, E., and Tomos, A.J.T.N.P. (1997). Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (Oryza sativa) **137,** 83-91.

Raven, J.A., and Edwards, D. (2004). Physiological evolution of lower embryophytes: adaptations to the terrestrial environment. In The evolution of plant physiology (Elsevier), pp. 17-41.

Raven, P.H., Evert, R.F., and Eichhorn, S.E. (2005). Biology of plants. (Macmillan).

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S.J.P.o.t.N.A.o.S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome **98,** 11479-11484.

Robinson, D.O., Roeder, A.H.J.C.O.i.G., and Development. (2015). Themes and variations in cell type patterning in the plant epidermis **32,** 55-65.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (Springer), pp. 234-241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., and Bernstein, M.J.I.j.o.c.v. (2015). Imagenet large scale visual recognition challenge **115,** 211-252.

Simonyan, K., and Zisserman, A.J.a.p.a. (2014). Very deep convolutional networks for large-scale image recognition.

Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol **6,** e1000770.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc **7,** 500-507.

Suh, M.C., and Go, Y.S. (2014). DEWAX-mediated transcriptional repression of cuticular wax biosynthesis in Arabidopsis thaliana. Plant Signal Behav **9,** e29463.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiol **139,** 1649-1665.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend, S.H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature **415,** 530-536.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am J Hum Genet **90,** 7-24.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet **101,** 5-22.

Wilson, A.G. (2014). Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes (University of Cambridge).

Xiang, J.-J., Zhang, G.-H., Qian, Q., and Xue, H.-W.J.P.p. (2012). Semi-rolled leaf1 encodes a putative glycosylphosphatidylinositol-anchored protein and modulates rice leaf rolling by regulating the formation of bulliform cells **159,** 1488-1500.

Yeats, T.H., and Rose, J.K.J.P.p. (2013). The formation and function of plant cuticles **163,** 5-20.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet **38,** 203-208.

Zeiler, M.D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (Springer), pp. 818-833.

Zhang, G.-H., Xu, Q., Zhu, X.-D., Qian, Q., and Xue, H.-W.J.T.P.C. (2009a). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development **21,** 719-735.

Zhang, G.H., Xu, Q., Zhu, X.D., Qian, Q., and Xue, H.W. (2009b). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. Plant Cell **21,** 719-735.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat Genet **42,** 355-360.

Zhao, S.Q., Hu, J., Guo, L.B., Qian, Q., and Xue, H.W. (2010). Rice leaf inclination2, a VIN3-like protein, regulates leaf angle through modulating cell division of the collar. Cell Res **20,** 935-947.

Zou, L.-p., Sun, X.-h., Zhang, Z.-g., Liu, P., Wu, J.-x., Tian, C.-j., Qiu, J.-l., and Lu, T.-g.J.P.p. (2011). Leaf rolling controlled by the homeodomain leucine zipper class IV gene Roc5 in rice **156,** 1589-1602.

Chapter 2: Network analyses implicate a role for PHYTOCHROMES and LIPID TRANSFER PROTEINS in the evolution of plant cuticles

2.1 Abstract

Plant cuticles are composed of wax and cutin, and evolved in the land plants as a hydrophobic boundary that reduces water loss from the plant epidermis. The expanding maize adult leaf displays a dynamic, proximodistal gradient of cuticle development, from the leaf base to the tip. Laser microdissection RNA Sequencing (LM-RNAseq) was performed along this proximodistal gradient, and complementary network analyses identified potential regulators of cuticle biosynthesis and deposition. Correlations between cuticle development and cell wall biosynthesis processes were identified, as well as evidence of roles for auxin and brassinosteroids. In addition, our network analyses suggested a previously undescribed function for PHYTOCHROME-mediated light signaling during cuticular wax deposition. Reverse genetic analyses confirmed that the *phyb1 phyb2* double mutant of maize and the *phy3* mutant of the moss *Physcomitrella patens* exhibit abnormal cuticle composition. These findings suggest a model for light-stimulated development of cuticular waxes during plant evolution. Transcriptomic analyses of light/dark exposed maize plants implicate a specific LIPID TRANSFER PROTEIN (LTP), which is activated by PHYB signaling, in the abnormal cuticle phenotype of *phyb1 phyb2* double mutants. LTPs and cuticles both evolved in the land plants, and are absent from aquatic green algae. We propose

that during plant evolution, LTPs arose as an innovation of land plants that enabled development of the cuticle.

2.2 Significance Statement

Plant cuticles provide barriers to water loss, and arose as aquatic plants adapted to the dry terrestrial environment. Cuticle waxes are synthesized in the plant epidermis, and deposited on the plant surface via transporters such as LIPID TRANSFER PROTEINS (LTPs). This study identifies a role for PHYTOCHROME light receptors during cuticle development in maize and in the moss plants. Mosses are remnants of early-diverging plant lineages, and are separated from flowering plants such as maize by over 400 million years. We reveal a role for PHYTOCHROME-mediated cuticle development in evolutionarily diverse plants, and show that a maize LTP is activated by PHYTOCHROME signaling. We propose that light-stimulated LTP function was critical to the evolution of cuticles in land plants.

2.3 Introduction

Light perception plays important roles in the regulation of plant metabolism and development (Tobin and Silverthorne, 1985; Chory and Susek, 1994; Fankhauser and Chory, 1997; Mittmann *et al.*, 2004; Whitelam and Halliday, 2007), including the activation of lipid production in algae (Sorigue *et al.*, 2016; Sorigue *et al.*, 2017). However, aquatic algae lack cuticles, the hydrophobic barrier deposited on the epidermis of all land plants that prevents nonstomatal water evaporation through the plant surface. Cuticles cover the above ground shoot of land plants, and enabled their invasion and colonization of the dry and hostile terrestrial environment. Since the

majority of water loss in plants occurs through the epidermis, the cuticle imparted a significant advantage in plant evolution by providing a barrier to desiccation (Kenrick and Crane, 1997b; Bateman *et al.*, 1998a; Raven and Edwards, 2004; Kerstiens, 2006a; Jetter and Riederer, 2016b).

The plant cuticle comprises a mixture of solvent-soluble lipids (waxes) plus a lipid polymer (cutin). Waxes are long-chain, non-polar molecules, composed mainly of hydrocarbons (alkanes and alkenes), aldehydes, alcohols, ketones and wax esters. In contrast, cutins are polymers of hydroxy fatty acids connected by ester bonds (Kolattukudy, 2001b; Nawrath, 2002a; Yeats and Rose, 2013b). Waxes and cutins are both formed *de novo* from long-chain C_{16} and C_{18} fatty acids synthesized within plastids of the plant epidermis (Samuels *et al.*, 2008; Kunst and Samuels, 2009a). In *Arabidopsis thaliana*, these long-chain fatty acids are converted to CoA thioesters by LONG-CHAIN ACYL-COENZYME A SYNTHASE (LACS), and subsequently transported into the endoplasmic reticulum, where they are elongated by the fatty acid elongase (FAE) complex (Joubès *et al.*, 2008). After elongation and further modification, these cuticle lipids are exported out of the plasma membrane and into the apoplastic space, where small protein transporters such as LTPs may facilitate the transport of lipid precursors to the site of cuticle deposition (Debono *et al.*, 2009; Lee *et al.*, 2009; Kim *et al.*, 2012). Cuticle development is regulated by many factors such as the phytohormone abscisic acid (ABA), water deficit, osmotic stress, and light (Hooker *et al.*, 2002; Joubès *et al.*, 2008; Suh and Go, 2014).

Previous studies in Arabidopsis have shown that the expression of several cuticle biosynthesis genes is induced by light (Hooker *et al.*, 2002; Joubès *et al.*, 2008; Suh

and Go, 2014). Intriguingly, light-activated photoenzymes can stimulate the enzymatic conversion of fatty acids to hydrocarbons in green algal relatives of land plants (Sorigue *et al.*, 2016; Sorigue *et al.*, 2017), although these lipids are not deposited on algal surfaces to form a cuticle. Likewise, PHYTOCHROME light receptors, which regulate a variety of physiological processes during plant growth and development, are also found in green algae (Duanmu *et al.*, 2014; Li *et al.*, 2015). LTPs, on the other hand, are only found in land plants, and are proposed to play a pivotal role in cuticle biosynthesis (Edstam *et al.*, 2011; Finkina *et al.*, 2016; Salminen *et al.*, 2016; Edqvist *et al.*, 2018; Salminen *et al.*, 2018).

In this study, we utilized the expanding adult leaf as a model system to elucidate the spatial-temporal gradient of maize cuticle development. Transcriptomic analyses were performed along the proximodistal axis of the developing maize leaf eight as it emerged from darkness to light. Complementary network analyses of transcriptomic data were employed to identify patterns of epidermal gene expression underlying the cuticle composition gradient previously identified in this tissue (Bourgault *et al.*, 2019a), and to identify novel candidate genes for cuticle development in maize. Network analyses suggested a previously unidentified role for PHYTOCHROME during cuticle development, which was confirmed by genetic and biochemical investigations in the evolutionarily divergent model plants *Zea mays* and *Physcomitrella patens*. We propose a mechanistic model whereby the innovation of LTP function in land plants was one critical step enabling the evolution of cuticle development.

2.4 Results and Discussion

2.4.1 Transcriptomic analyses of cuticle development in the adult maize leaf

Previous analyses demonstrated a gradient of cuticle composition along the proximodistal axis of the expanding leaf eight of maize inbred line B73, from light-shielded proximal intervals to light-exposed distal regions (Figure 2.1A). In general, longer-chain wax components and cutin monomers increase in abundance as the leaf transitions from the dark to light (Figure 2.1B) (Bourgault *et al.*, 2019a). To capture the transcriptional gradient coinciding with these biochemical changes in cuticle composition, plants were grown under equivalent environmental conditions as in (Bourgault *et al.*, 2019a), and seven developmental stages along the proximodistal axis of leaf eight were laser-microdissected for RNAseq analysis. Each stage comprised a two-centimeter-long interval, collected between 2 and 22 cm from the leaf base, encompassing the point of emergence of leaf tissue into the light at ~18 cm (Figure 2.1A). For each of the seven proximodistal intervals examined, an L1-derived epidermal sample and an L2-derived internal sample were microdissected (Figure 2.2), followed by RNA sequencing to construct their transcriptomes. Principal component analysis (PCA) identified two PCs that collectively explain 60.29% of the total sample variance in the transcriptomic data. Specifically, the first PC explains 38.22% of the total sample variation, and corresponds to the seven proximodistal leaf intervals analyzed, whereas the second PC (PC2, 22.07% of sample variance) delineates epidermal and internal tissues for each leaf developmental stage (Figure 2.3). These data show that in addition to the biochemical gradient in cuticle composition, the leaf intervals examined in maize leaf 8 also exhibit a transcriptomic gradient.

2.4.2 Identification of putative cuticle regulatory genes across the leaf developmental gradient via directed network inference

A gene regulatory network (GRN) was constructed using causal structure inference (CSI), as a means to identify candidate regulators of cuticle biogenesis based on analysis of the epidermal transcriptomic patterns in the seven intervals of maize leaf eight. The CSI algorithm accounts for the biological delay of one gene's effect on another gene's expression level, by incorporating both the spatial and temporal transcriptomic information within all the sampled leaf intervals (Penfold and Wild, 2011). CSI generates a network with connections (edges) between genes (nodes), directed from putative regulators toward their targets. A Bonferroni-corrected significance level was used to identify putative regulatory relationships among epidermally-enriched transcripts from the seven leaf intervals, since the epidermis is the site of cuticle biosynthesis (Suh *et al.*, 2005b; Kunst and Samuels, 2009a).

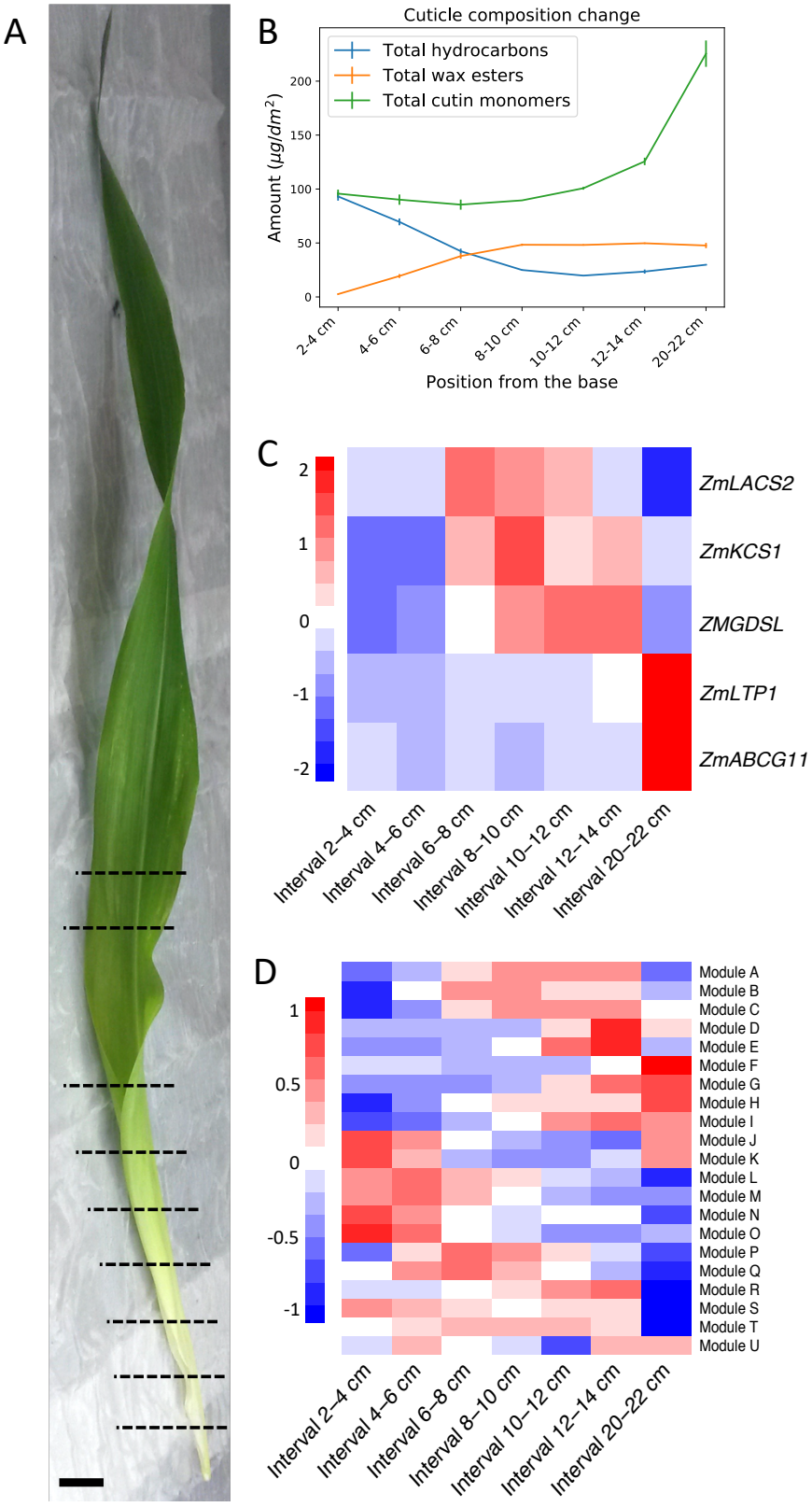
A
B
Cuticle composition change
Total hydrocarbons
Total wax esters
Total cutin monomers
Amount (μg/dm²)
200
150
100
50
0
2-4 cm
4-6 cm
6-8 cm
8-10 cm
10-12 cm
12-14 cm
20-22 cm
Position from the base
C
2
1
0
-1
-2
ZmLACS2
ZmKCS1
ZMGDSL
ZmLTP1
ZmABCG11
Interval 2–4 cm
Interval 4–6 cm
Interval 6–8 cm
Interval 8–10 cm
Interval 10–12 cm
Interval 12–14 cm
Interval 20–22 cm
D
1
0.5
0
-0.5
-1
Module A
Module B
Module C
Module D
Module E
Module F
Module G
Module H
Module I
Module J
Module K
Module L
Module M
Module N
Module O
Module P
Module Q
Module R
Module S
Module T
Module U
Interval 2–4 cm
Interval 4–6 cm
Interval 6–8 cm
Interval 8–10 cm
Interval 10–12 cm
Interval 12–14 cm
Interval 20–22 cm

Figure 2.1. Proximodistal transcriptomic gradients in the expanding adult maize leaf. (A) Dashed lines demarcate the boundaries of 2-cm intervals collected along the proximodistal axis of the expanding maize leaf eight (note: the leaf emerges from the whorl into the light ~18 cm from the leaf base). Scale bar (solid line) = 2 cm. (B) Line plots showing changes in cuticle component abundance along the proximodistal axis of leaf 8. Proximal leaf regions accumulate more hydrocarbons (alkanes and alkenes), whereas distal regions contain higher levels of wax esters and cutin monomers. (C) Heatmap showing the expression patterns of selected cuticle biosynthesis and transport genes across seven proximodistal intervals of leaf 8. (D) Heatmap showing the expression levels of eigengenes for the 21 modules identified in our gene co-expression network.

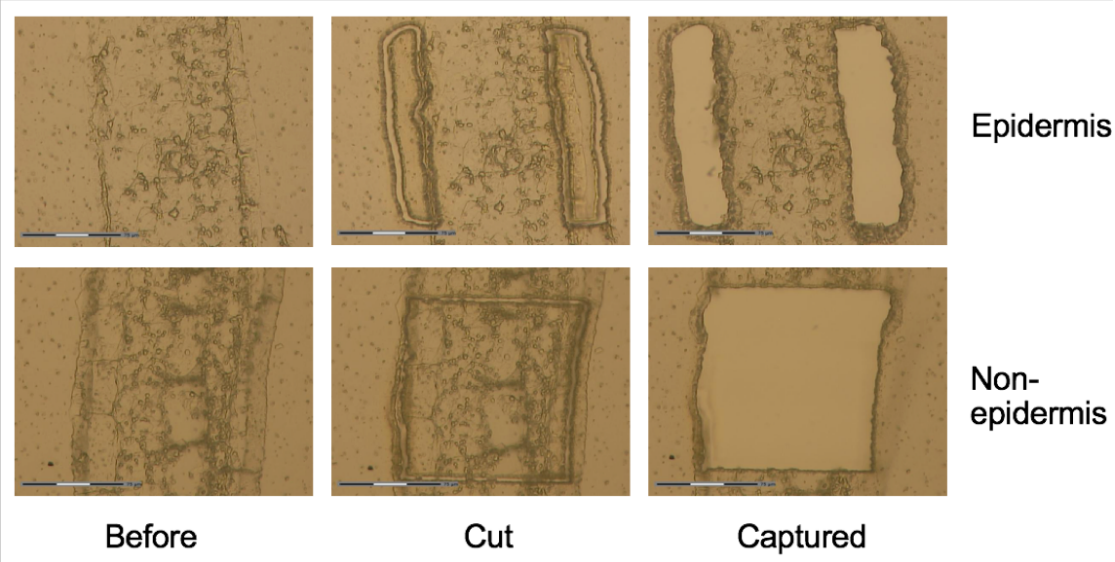

Figure 2.2. Laser microdissection was performed on leaf tissues to isolate the L1-derived epidermal layers (top row) and L2-derived internal layers of each targeted leaf interval along the proximodistal axis of the expanding leaf 8. From left to right, each column corresponds to a leaf section before, during and after microdissection. Scale bars = 75 μm.

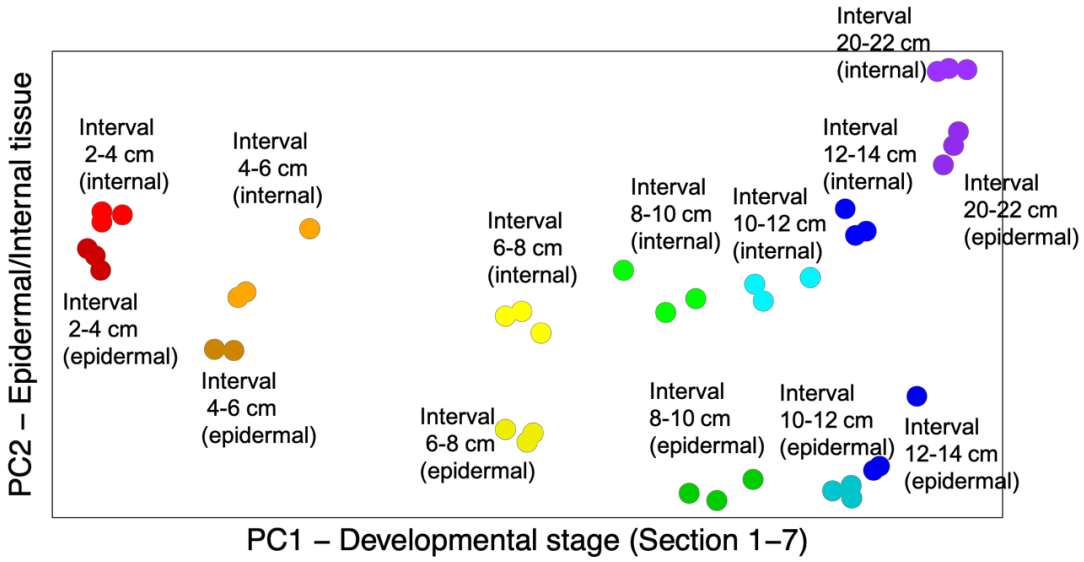

Figure 2.3. Principal component analysis identified two principal components (PCs) corresponding to the developmental stage (PC1) and tissue type (PC2) of leaf 8 samples in our LM-RNAseq analysis. Each point corresponds to one RNAseq sample. From left to right, each color corresponds to a specific developmental stage (youngest to oldest part of the leaf). From bottom to top, darker color shades represent L1-derived epidermal tissues, and lighter shades represent L2-derived internal tissues. For interval two, one outlier epidermal sample was removed.

Previously described cuticle biosynthetic genes were differentially expressed across the seven leaf intervals (Figure 2.1B), coinciding with the changes in cuticle composition in the expanding leaf eight. For example, expression levels of ABCG transporters (e.g. *ZmABCG11* in Figure 2.1B) that deliver wax components and cutin monomers out of the plasma membrane (Pighin *et al.*, 2004; Bird *et al.*, 2007; McFarlane *et al.*, 2010) show a consistent increase that correlates with the accumulation of cutin monomers (Bourgault *et al.*, 2019a) within these intervals. These data established that there is a gradient in the epidermal gene expression of leaf 8, especially for transcripts involved in cuticle biosynthesis and deposition. The gene expression gradients can be incorporated into the GRN model as the "biological delay" of their regulators' functions, supporting the efficacy of our CSI approach to utilize the regulatory delay. Known cuticle biosynthetic genes were used as "baits" to identify putative regulators of cuticle biogenesis. Baits included *ECERIFERUM3* (Rowland *et al.*, 2007) (*GRMZM2G029912*) and *BETA-KETO-ACYL REDUCTASE* (*KCR*) (Beaudoin *et al.*, 2009) (*GRMZM2G090733*), and are highlighted in yellow in Figure 2.4. Known cuticle biosynthesis nodes showed incoming edges from genes with no previously described function in cuticle development (such as *GRMZM2G055469*, *GRMZM2G120619*, *GRMZM2G078959*) (Table S1), identifying these new genes as potential regulators of cuticle biosynthesis. Although many of these candidate genes are not previously described to function during the regulation of cuticle development, others such as *GRMZM2G055158*, have homologs in Arabidopsis (*MYB20*) that are involved in secondary cell wall thickening (Zhong *et al.*, 2008). Thus, these genes with numerous outgoing edges, represent potential candidate genes for reverse genetic analyses of maize cuticle biosynthesis.

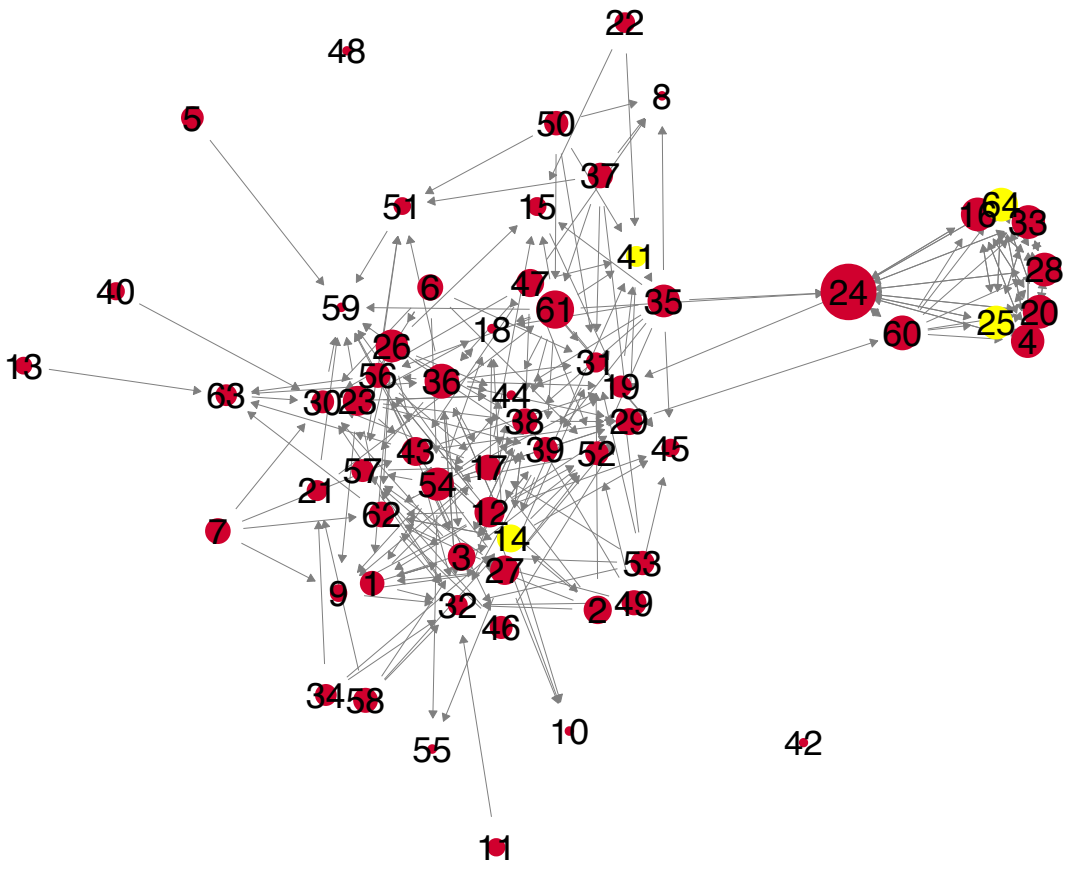

Figure 2.4. Gene regulatory network of epidermally enriched genes in the expanding adult leaf 8 of maize. Previously described cuticle regulatory genes, and genes functioning in cuticle biosynthesis or lipid transport and are highlighted in yellow. Arrows point to potential targets of the corresponding regulatory genes. The size of the nodes correlates with the number of targets for each node (directional edges coming out of the node), i.e. outdegree. Each number corresponds to a specific gene identified.

2.4.3 A weighted co-expression network analysis identifies additional candidate regulatory genes for cuticle biosynthesis

Although the CSI algorithm is a powerful tool enabling gene discovery and their functional relationships, its ability to model the biological delay of a gene's effect can also be a disadvantage, depending upon how the transcriptomic samples are collected. For example, although CSI will detect the delayed effects of a gene *between* two 2 cm leaf intervals, it will fail to detect these effects *within* any 2 cm leaf sample. Hence, we utilized a second, complementary method called gene co-expression network (GCN) analysis (Langfelder and Horvath, 2008, 2012) to identify additional candidate genes involved in regulating the biosynthesis of the maize cuticle. GCN is essentially a "guilt-by-association" approach, wherein correlations in gene expression levels implicate co-regulation of gene pairs within the network.

We constructed a weighted gene co-expression network based upon the expression-level correlations of all 11,816 epidermally-transcribed genes identified in our LM-RNAseq (Langfelder and Horvath, 2008). The transcriptome was partitioned into 21 modules; Figure 2.1C illustrates the expression levels of eigengenes (idealized representative genes) within these 21 modules at each of the seven, leaf developmental stages analyzed. Several developmental trends are obvious in these data. For example modules L – O exhibit a gradient of decreasing transcript abundance moving from proximal to distal leaf intervals, while modules F- I show the opposite trend. Comparisons of gene expression levels to cuticle lipid profiles at each interval reveal interesting correlations, as shown in Figure 2.5. Specifically, the expression levels of genes in modules C, H and I are positively correlated with the accumulation of wax

esters, whereas modules L, M, N and O are negatively correlated with these wax components. Modules F – I also correlate with increasing abundance of cutin monomers and very long chain aliphatics (C_{31} – C_{35}). Modules C, F, I, Q (Tables S4, S7, S10, S18) showing positive or negative correlations with specific cuticle lipid profiles also contain known cuticle biosynthesis and regulatory genes; for example, in module F (Table S7), the *KCS6* homologs *AC233893.1_FG003* and *GRMZM2G060481* show correlation coefficients of 0.95 and 0.81, respectively, with the accumulation of alkane C_{35}. The strategic use of this network for discovery of candidate genes important for cuticle development is described below.

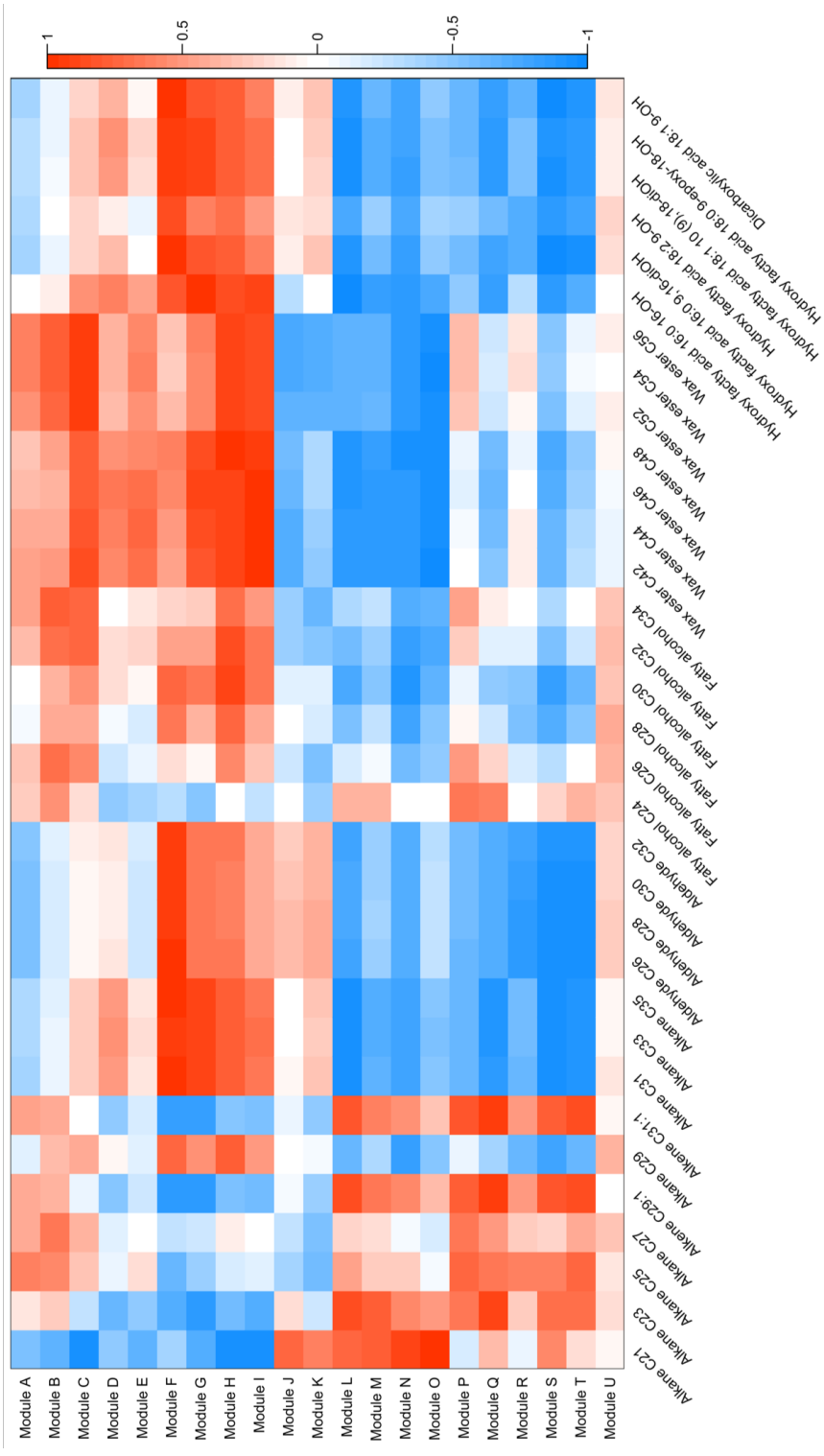
1
0.5
0
-0.5
-1
Module A
Module B
Module C
Module D
Module E
Module F
Module G
Module H
Module I
Module J
Module K
Module L
Module M
Module N
Module O
Module P
Module Q
Module R
Module S
Module T
Module U
Alkane C21
Alkane C23
Alkane C25
Alkane C27
Alkane C29:1
Alkane C29
Alkane C31:1
Alkane C31
Alkane C33
Alkane C35
Aldehyde C26
Aldehyde C28
Aldehyde C30
Aldehyde C32
Fatty alcohol C24
Fatty alcohol C26
Fatty alcohol C28
Fatty alcohol C30
Fatty alcohol C32
Fatty alcohol C34
Wax ester C42
Wax ester C44
Wax ester C46
Wax ester C48
Wax ester C52
Wax ester C54
Wax ester C56
Hydroxy fatty acid 16:0 16-OH
Hydroxy fatty acid 16:0 9,16-diOH
Hydroxy fatty acid 18:1 10 (9),18-diOH
Hydroxy fatty acid 18:2 9-OH
Dicarboxylic acid 18:0 9-epoxy-18-OH
Hydroxy fatty acid 18:1 9-OH

Figure 2.5. Heat map depicting the correlation of each cuticle lipid component (x axis) with the 21 co-expression modules (y axis) identified in transcriptomic analyses of the expanding maize leaf 8. Colors (red to blue) correspond to the values of the Pearson's pairwise correlations, wherein red (+1) is positively correlated and blue (-1) is negatively correlated. Cuticle lipid abundance data used for this analysis are described in (Bourgault *et al.*, 2019a).

One such strategy is to interrogate the direct neighbors of known cuticle biosynthetic genes within the network, as demonstrated in analyses of five modules (Q, F, C, A and I) showing enrichment for cuticle genes. In particular, members of several gene families known to function in cuticle biosynthesis and transport are overrepresented in Module F. These include *KCS*s, *ABC TRANSPORTER*s, and *LTP*s (Table S7) (Pighin *et al.*, 2004; Joubès *et al.*, 2008; Debono *et al.*, 2009), whose direct neighbors exhibit strong co-expression with these cuticle genes, indicating regulatory roles or co-regulation. The direct neighbors of known cuticle biosynthesis genes in these modules comprise additional potential candidate genes involved in the regulation of cuticle biosynthesis in the expanding maize leaf (Langfelder and Horvath, 2008; Vandepoele *et al.*, 2009). Intriguingly, epidermally-enriched transcripts are overrepresented in these modules. For example, 41.26% of the 972 genes in Module F are upregulated in the epidermis in comparison to internal tissues, whereas only 31.89% of all the 11,816 epidermally-transcribed genes are upregulated in the epidermis. This additional layer of spatial filtering further support the associations between these modules and cuticle biogenesis.

A second gene discovery approach enabled by GCN analyses is to examine the hubs within the network, defined as the most connected nodes that are essential to network function (Barabasi and Oltvai, 2004; Barabasi, 2013). In our case, the hubs of modules that are significantly correlated with cuticle components are usually a subset of the candidates identified via the "direct neighbor" approach described above. For example, in module F many of the hubs are also direct neighbors of homologs of genes encoding KCSs, ABC TRANSPORTERs, and LTPs. These findings further

confirm the importance of module F, and especially its hubs, on cuticle development in the expanding maize leaf.

2.4.4 Light-regulated cuticle development: *phytochrome* (*phy*) mutants have altered cuticle composition

Previous studies showed that light induces cuticular wax biosynthesis in land plants and that the expression levels of several fatty acid elongase complex transcripts decrease in dark-grown plants, thus reducing the amount of cuticular wax (Hooker *et al.*, 2002; Joubès *et al.*, 2008; Suh and Go, 2014; Kim *et al.*, 2018). Moreover, biochemical analyses also revealed that longer chain wax components are more abundant in the distal, light-exposed intervals of maize leaf 8 (Figure 2.1A; (Bourgault *et al.*, 2019a). Although algal relatives of the land plants do not develop a cuticle, light exposure does induce the production of hydrocarbons from long-chain fatty acids (Sorigue *et al.*, 2016; Sorigue *et al.*, 2017).

A GO term analysis of module H showed significant enrichment of transcripts that respond to light stimuli, and our GCN identified a previously-undescribed correlation between *PHYTOCHROMEs* (*PHYs)* and cuticle development. PHYs are red/far-red light photo-reversible chromoproteins that regulate gene expression in response to light. Maize contains six *PHY* homologs (*PHYA1*, *PHYA2, PHYB1, PHYB2, PHYC1, PHYC2*) . *PHYB1* and *PHYA1* occupy central positions in module H, whereas *PHYA2* and *PHYC* are peripheral nodes in the network (Figure 2.6A). Module H is also strongly correlated with many cuticle components, such as C_{31} - C_{35} alkanes (Figure 2.5). Gas chromatography–mass spectrometry (GC-MS) comparisons were performed on tissues derived from the blade (i.e. midrib and margins removed) portions of flag

leaves harvested from the previously described *phyb1 phyb2* double mutant and non-mutant siblings (Sheehan *et al.*, 2007). *phyb1 phyb2* double mutants show pleiotropic phenotypes in shoot development, including early flowering time, reduced internode width, and thinner leaves, although the *phyb1 phyb2* mutant flag leaf blades sampled in this analysis had otherwise equivalent morphological development as their non-mutant siblings. For example, the midrib, marginal, vasculature, and epidermal patterning phenotypes of *phyb1 phyb2* and non-mutant sibling flag leaves are indistinguishable (Sheehan *et al.*, 2007).

The *phyb* double mutant cuticles showed significant differences (at 5% significance level) for several cuticle wax components. For example, decreased amounts of alkanes and fatty alcohols were observed for molecules of 32 carbons and longer; the C_{34} aldehydes were also reduced, whereas the wax esters were unchanged. Although significant, the reductions in the amounts of longer-chain alkanes, fatty alcohols and aldehydes were relatively low (up to 4.3% of the total wax load in non-mutant siblings), and were accompanied by larger increases in shorter-chain components including: C_{24} - C_{30} fatty alcohols (Figure 2.6B), C_{23} - C_{31} alkanes, C_{28} - C_{32} aldehydes, C_{41}, C_{43}, C_{45}, C_{47}, C_{48}, and C_{50} wax esters (Figure 2.7 – 2.9) when compared with sibling leaves. We speculate that the increases in shorter-chain components are a compensatory response to the double mutant reduction in longer-chain compounds. These data strongly support models wherein PHY-mediated light signaling regulates cuticle biosynthesis, as predicted by module H. Interestingly, accumulation of transcripts encoding fatty acid elongase components such as *GRMZM2G164974* (*ZmCER6*) and *GRMZM2G104626* (*ZmKCS1*) are upregulated in the *phyb1 phyb2* double mutant, which stands in apparent contradiction to their reduction in long-chain

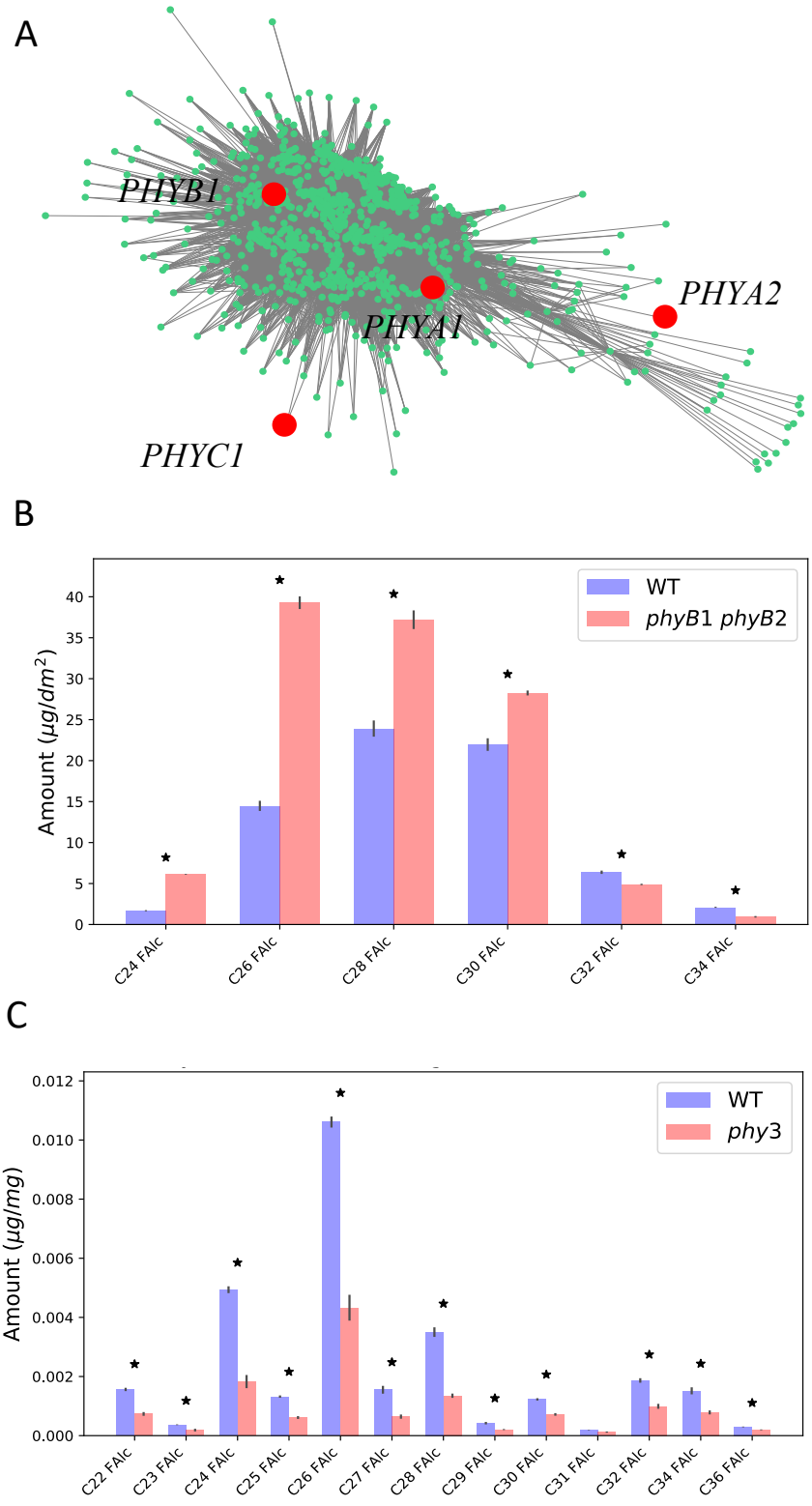
A
PHYB1
PHYA1
PHYA2
PHYC1
B
WT
phyB1 phyB2
Amount (μg/dm²)
C24 FAlc
C26 FAlc
C28 FAlc
C30 FAlc
C32 FAlc
C34 FAlc
C
WT
phy3
Amount (μg/mg)
C22 FAlc
C23 FAlc
C24 FAlc
C25 FAlc
C26 FAlc
C27 FAlc
C28 FAlc
C29 FAlc
C30 FAlc
C31 FAlc
C32 FAlc
C34 FAlc
C36 FAlc

Figure 2.6. Phytochrome regulates cuticle wax composition in the adult maize leaf 8. (A) Visualization of the co-expression network of module H. The red colored nodes correspond to the phytochrome homologs *PHYA1, PHYA2, PHYB1*, and *PHYC1*, some of which (*PHYB1* and *PHYA1*) occupy central positions of the network (with numerous connections with other nodes). Module H is also significantly correlated with specific cuticle components (Figure 2.5). (B) Comparisons of the fatty alcohol cuticular wax profiles in leaves of *phyB1 phyB2* double mutant versus nonmutant maize. (C) Comparisons of the fatty alcohols in cuticular wax profiles in gametophyte colonies of *phy3* mutant versus non-mutant *Physcomitrella patens*. Error bars represent standard errors. Asterisks indicate significant differences (< 0.05 p-value) between samples in unpaired t-tests.

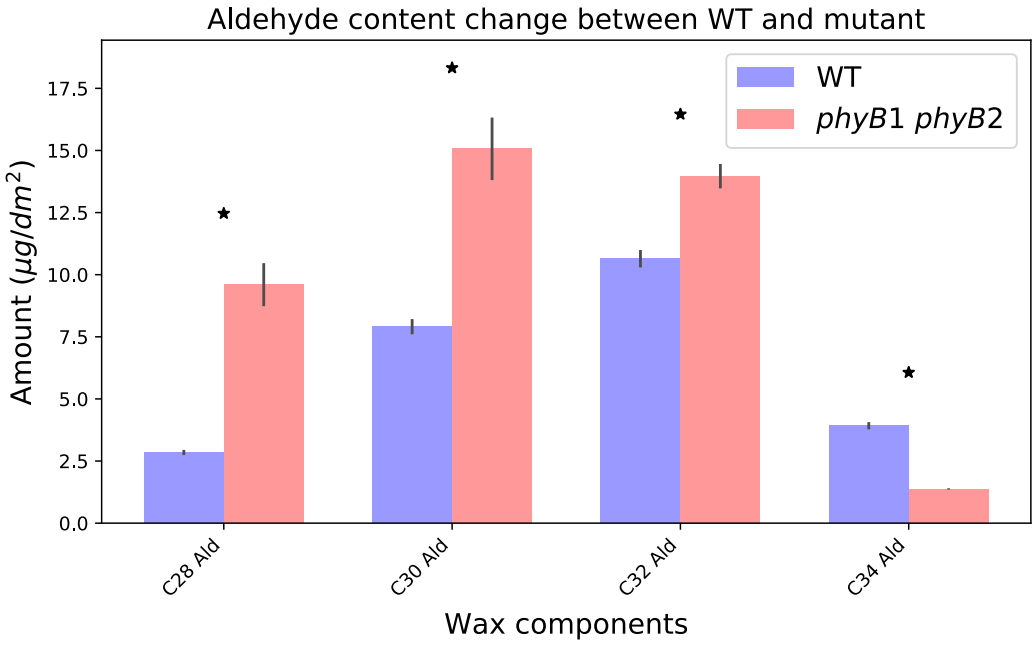

Figure 2.7. Comparisons of aldehydes in cuticular wax profiles in phyB1 phyB2 double mutants versus nonmutant plants in maize. Asterisks indicate significant differences (< 0.05 p-value) between samples. Error bars represent standard errors.

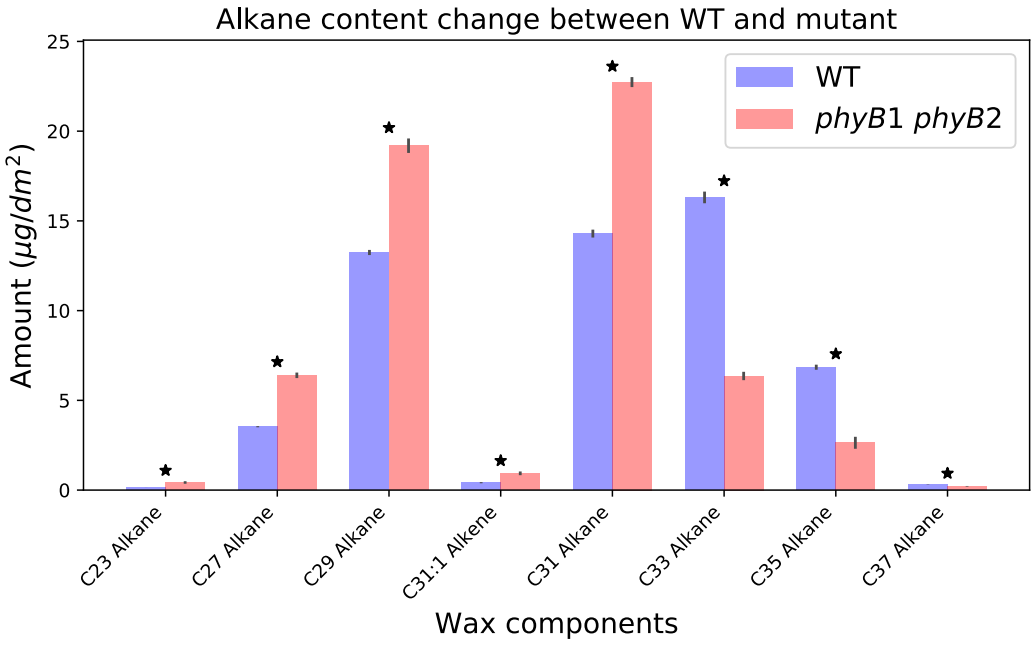

Figure 2.8. Comparisons of alkanes in cuticular wax profiles in phyB1 phyB2 double mutants versus nonmutant plants in maize. Asterisks indicate significant differences (< 0.05 p-value) between samples. Error bars represent standard errors.

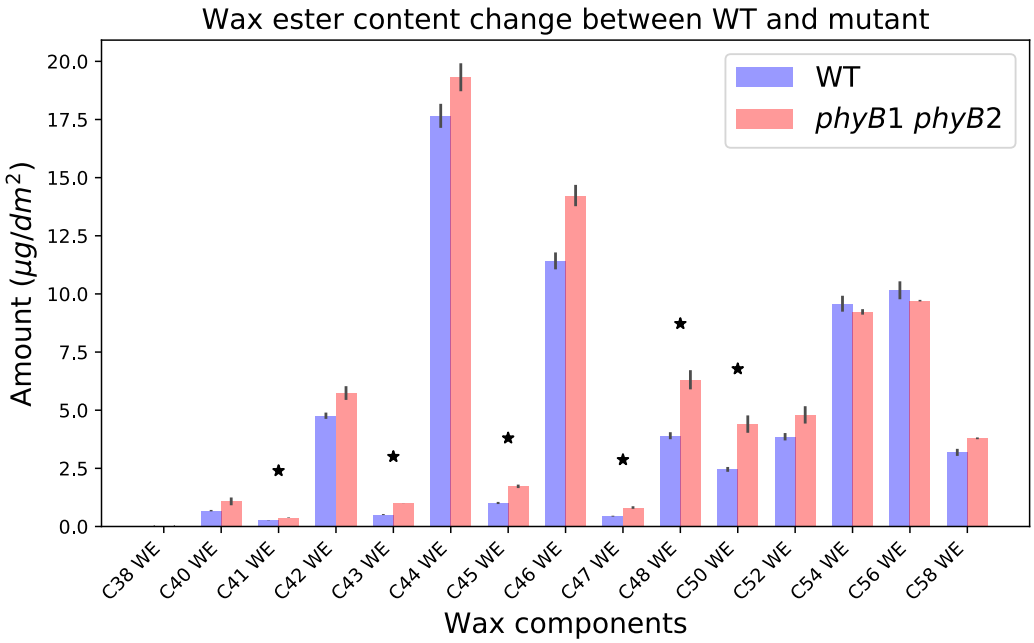

Figure 2.9. Comparisons of wax esters in cuticular wax profiles in phyB1 phyB2 double mutants versus nonmutant plants in maize. Asterisks indicate significant differences (< 0.05 p-value) between samples. Error bars represent standard errors.

wax components. However, the fatty acid elongase complex functions during biosynthesis of wax components with chain length < 32C (which are upregulated in *phyb1 phyb2* mutants) *as well as* wax compounds with chain length > 32C. Moreover, the reduction of wax compounds with chain length > 32C in the double mutant may not necessarily arise from a deficit in longer-chain biosynthesis, but rather from defects in the exportation of these wax components from epidermal cells. This hypothesis is further explored in RNAseq analyses described later.

Interestingly, the defects in cuticle composition observed in the *phyb1 phyb2* double mutant partially mirror the changes in cuticle components as the leaf emerges from the whorl (Bourgault *et al.*, 2019a). Specifically, the expression level of *PHYB1* is positively correlated with the accumulation of wax esters, and alkanes longer than C_{31}, both of which are more abundant in cuticles of distal sections of leaf 8 that are exposed to light. In contrast, *PHYB1* expression is negatively correlated with the accumulation of shorter chain alkanes, which are more abundant in proximal cuticle domains where the leaf is not exposed to light. Longer-chain wax components are increasingly abundant as leaf 8 is exposed to light (Bourgault *et al.*, 2019a), while the *phyB* double mutant shows defects in cuticle wax components that require fatty acid chain elongation beyond 32 carbons.

To determine whether PHY regulation of cuticle accumulation is simply a maize-specific phenomenon or is in fact found in other land plants, equivalent analyses of cuticle lipids were performed on *phy1, phy2, phy3* and *phy4* mutant colonies of the moss *Physcomitrella patens* (Mittmann *et al.*, 2004), a member of the bryophytes that diverged from later-evolved plant lineages early in the evolution of land plants

(Salminen *et al.*, 2018). *P. patens* encodes four PHY homologs (PHY1, PHY2, PHY3, PHY4), although none contain the canonical N-terminal extension that defines PHYB homologs (Mittmann *et al.*, 2004). All four *phy* mutants of *P.* patens display defects in cuticle wax components (Figure 2.6C, Figure 2.10 – 2.16). In *phy3* mutants, we observed reductions (at 5% significance level) in the amounts of all but one of the identified fatty alcohols (Figure 2.6C), in C_{29}, C_{31}, and C_{33} alkanes, in C_{25}, C_{26}, C_{27}, C_{28} and C_{30} aldehydes, and in C_{38}, C_{40}, C_{41}, C_{42}, C_{43}, C_{44}, and C_{45} wax esters (Figure 2.10 – 2.12) when compared with wild-type plants. The cuticle phenotype of the *phy3* mutant in moss is not equivalent to that of maize *phb1 phyb2*. However, the four *phy* mutants of moss do have complementary phenotypes (Figure 2.10 – 2.19). *phy3* is defective in longer-chain (carbon number ≥ 25) aldehydes, while *phy4* is defective in shorter-chain (carbon number ≤ 26) aldehyde components. Additionally, *phy2* is defective in aldehydes with chain length of 24C and 25C, but not the components longer or shorter than this range. These *phy* mutant moss phenotypes suggest that PHY regulation of cuticle development may be a widespread phenomenon of land plants, although analyses of additional plant taxa are required to test this model.

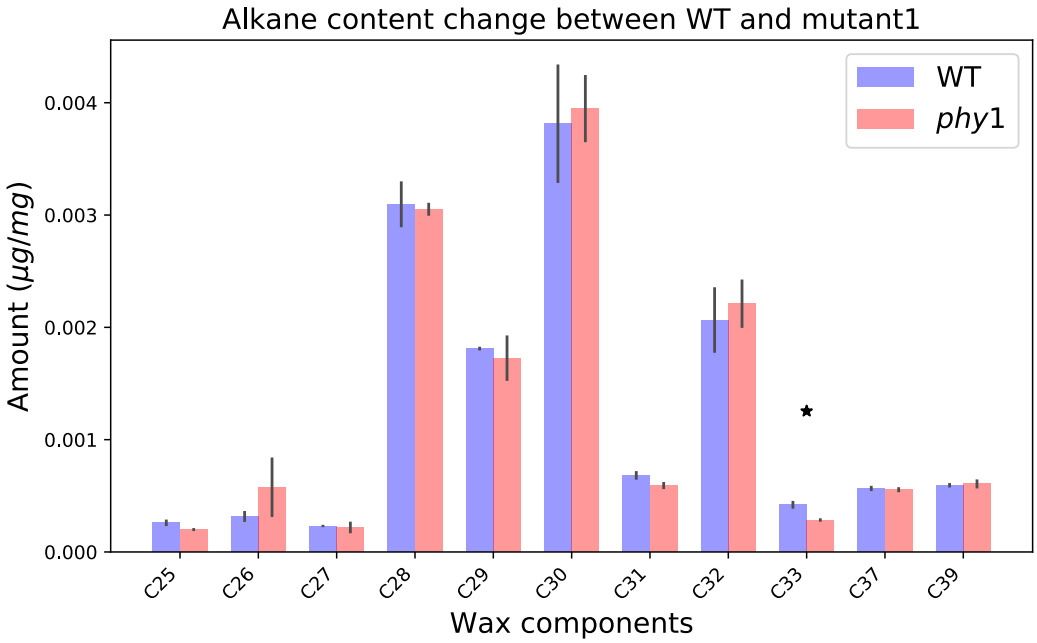

Figure 2.10. Comparisons of alkanes in cuticular wax profiles in *phy1* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

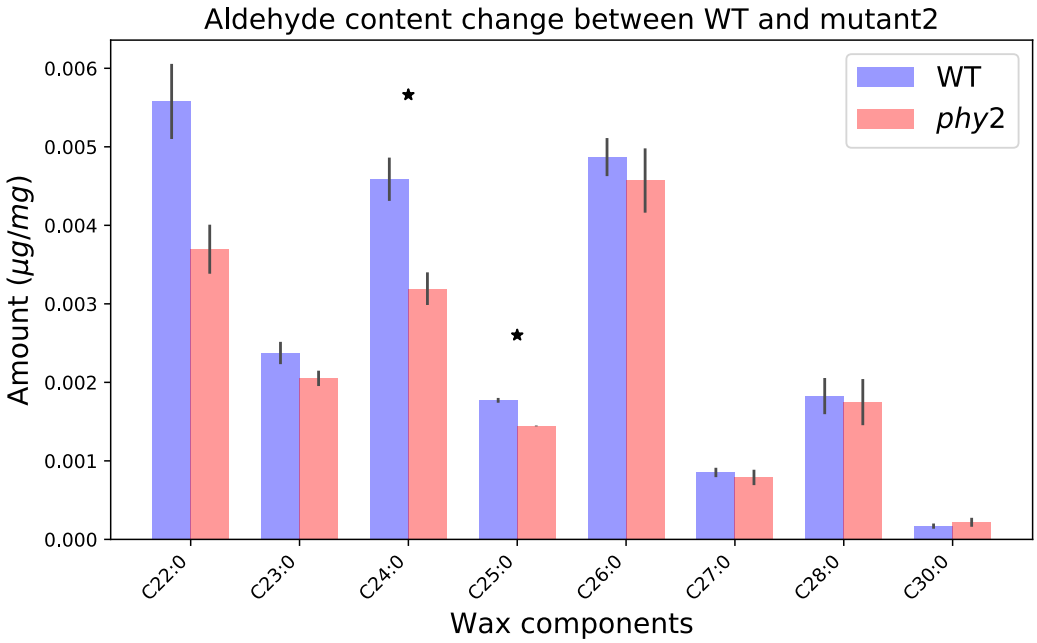

Figure 2.11. Comparisons of aldehydes in cuticular wax profiles in *phy2* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

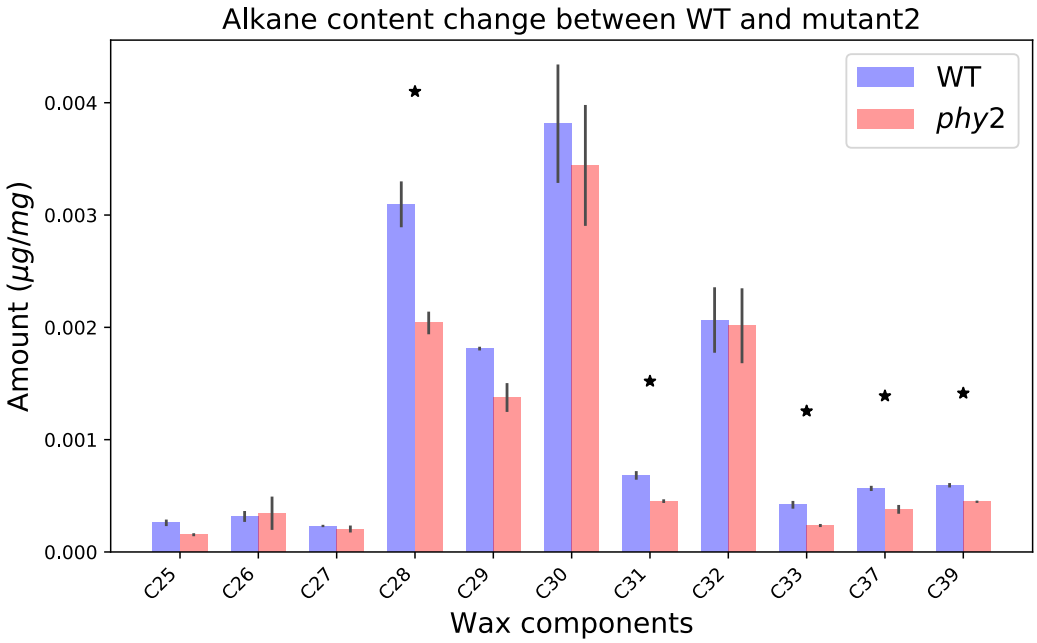

Figure 2.12. Comparisons of alkanes in cuticular wax profiles in *phy2* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

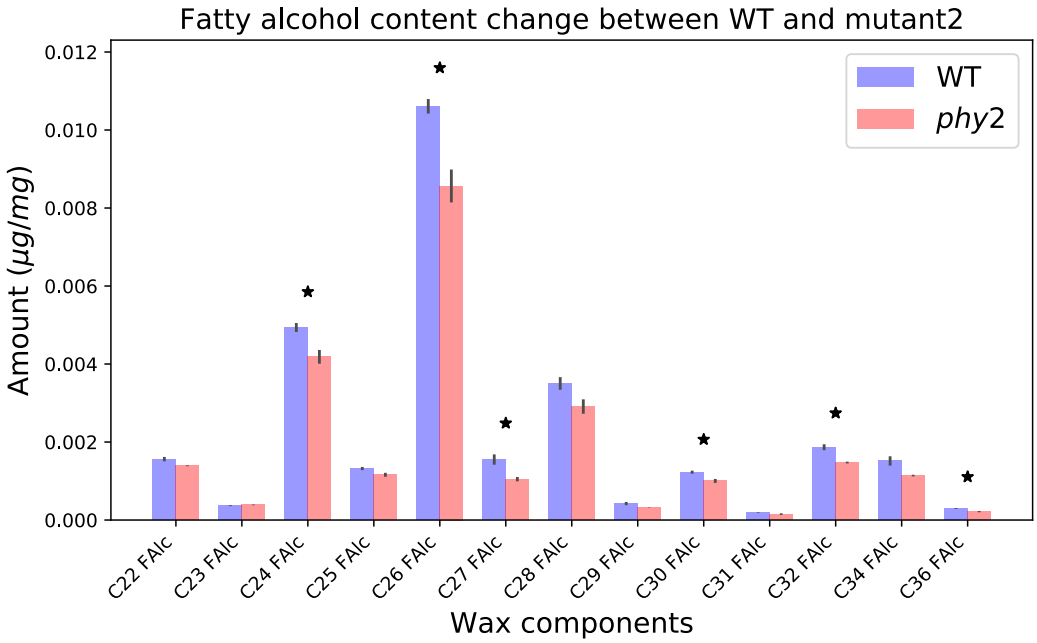

Figure 2.13. Comparisons of fatty alcohols in cuticular wax profiles in *phy2* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

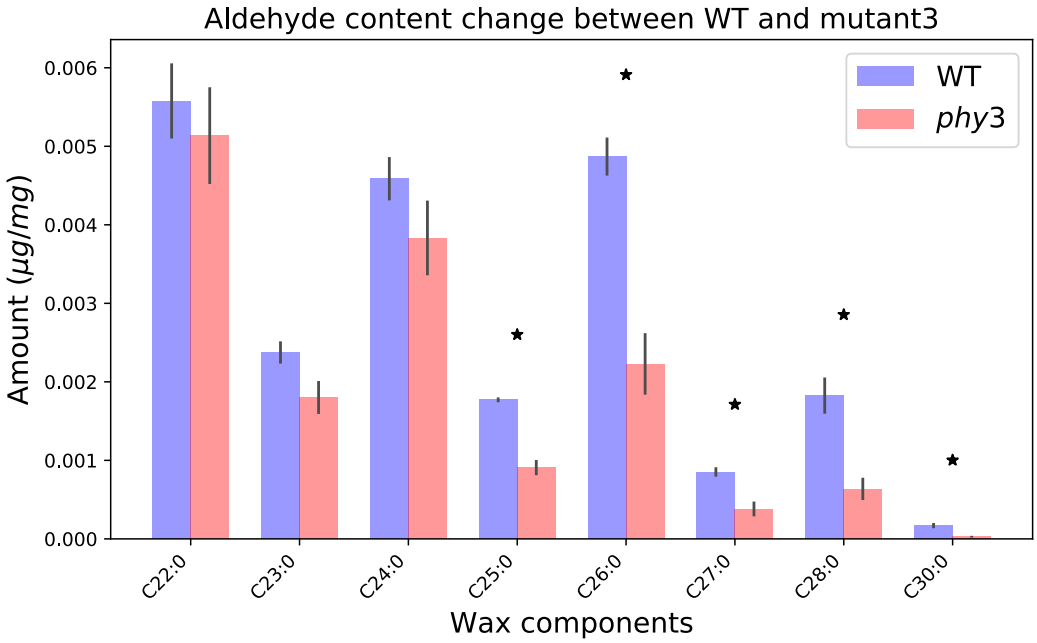

Figure 2.14. Comparisons of aldehydes in cuticular wax profiles in *phy3* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

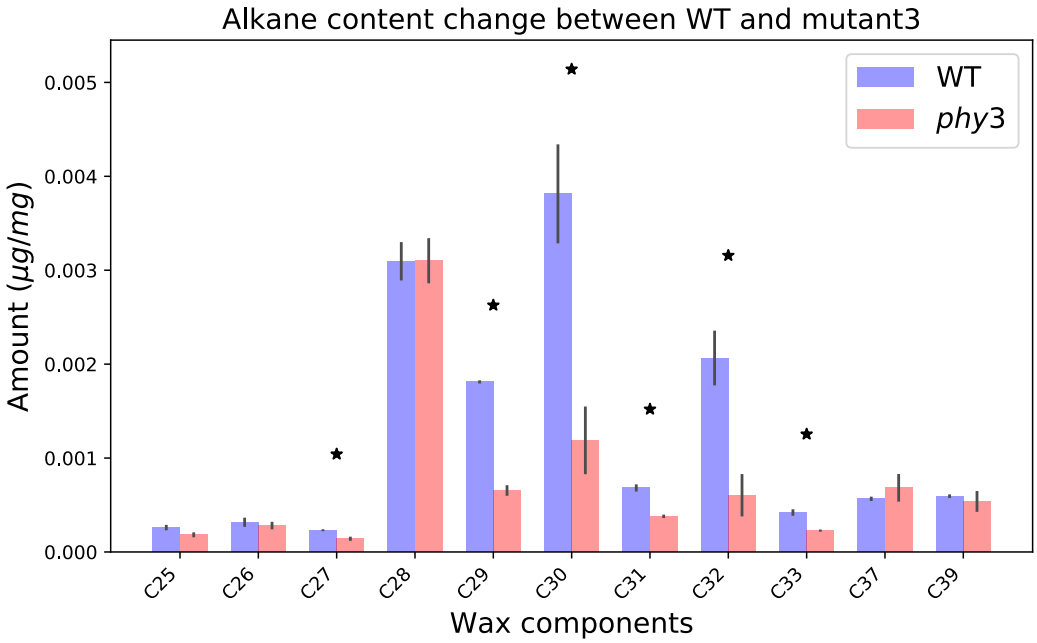

Figure 2.15. Comparisons of alkanes in cuticular wax profiles in *phy3* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

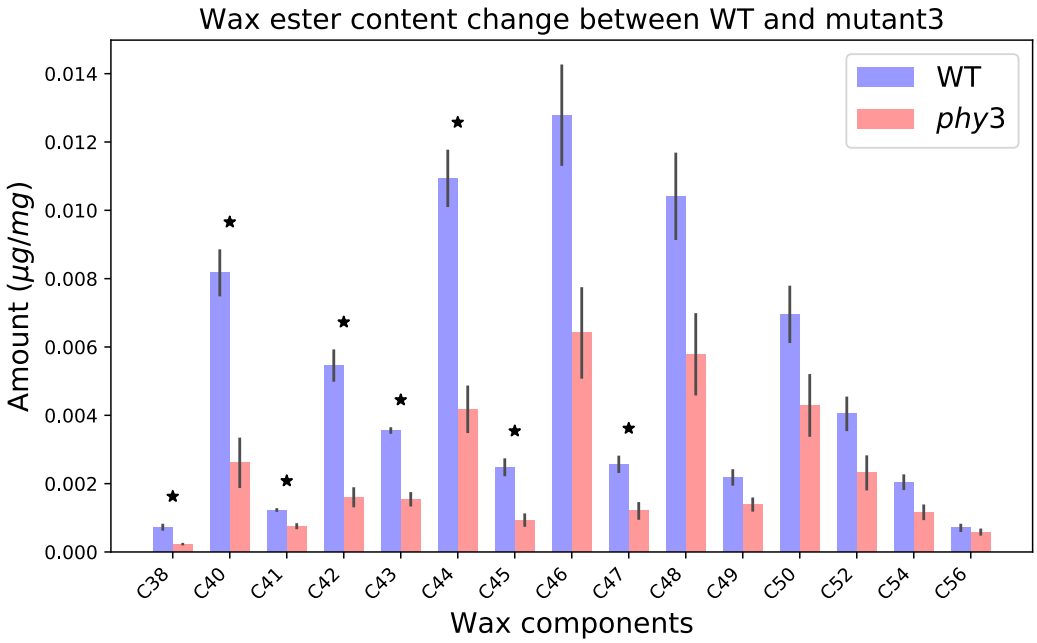

Figure 2.16. Comparisons of wax esters in cuticular wax profiles in *phy3* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

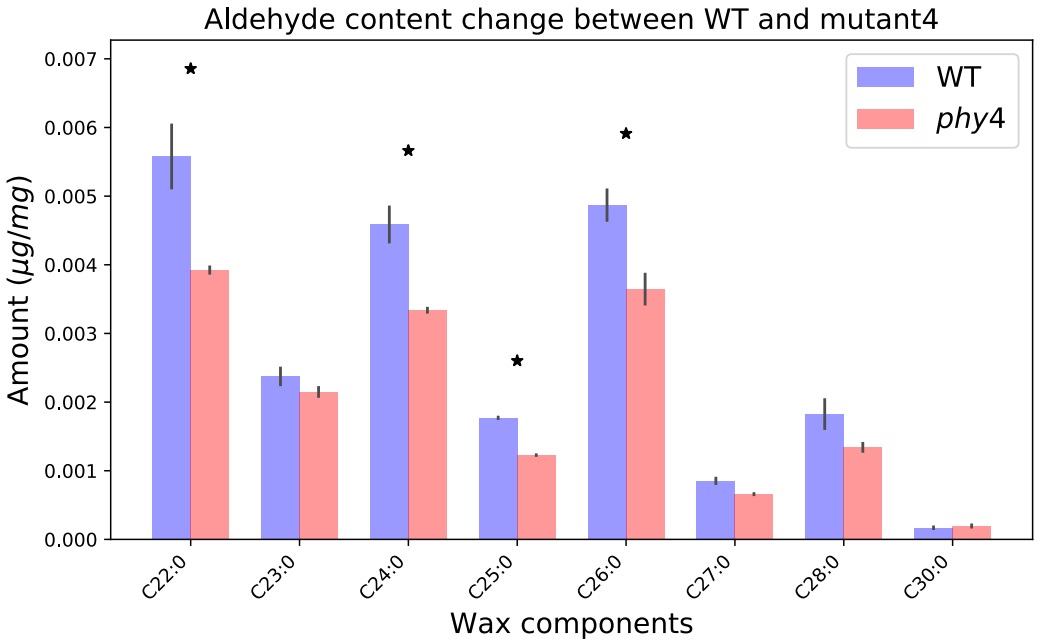

Figure 2.17. Comparisons of aldehydes in cuticular wax profiles in *phy4* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

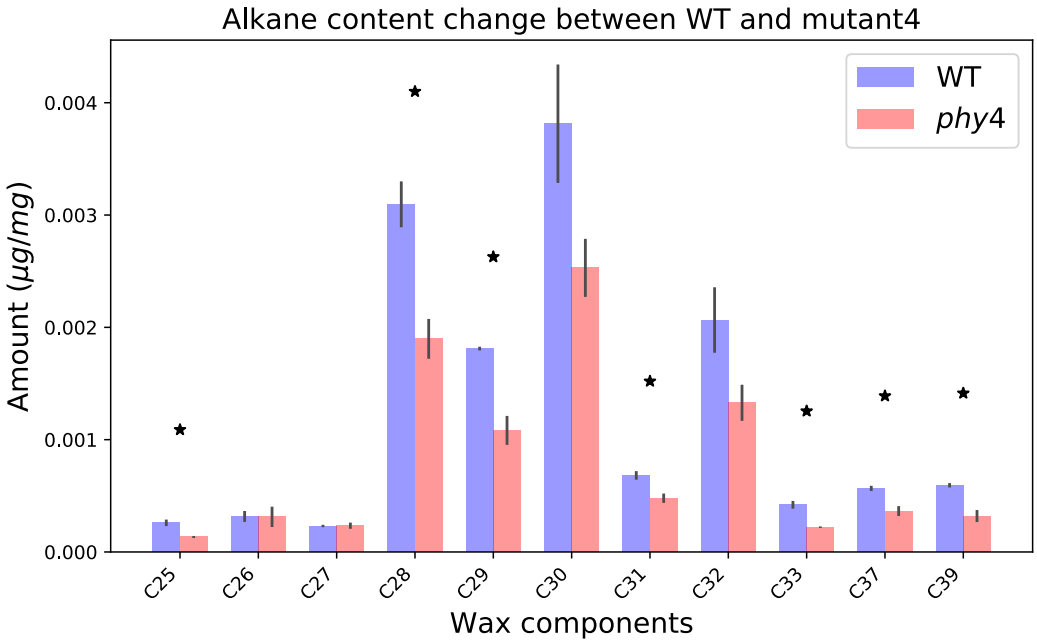

Figure 2.18. Comparisons of alkanes in cuticular wax profiles in *phy4* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

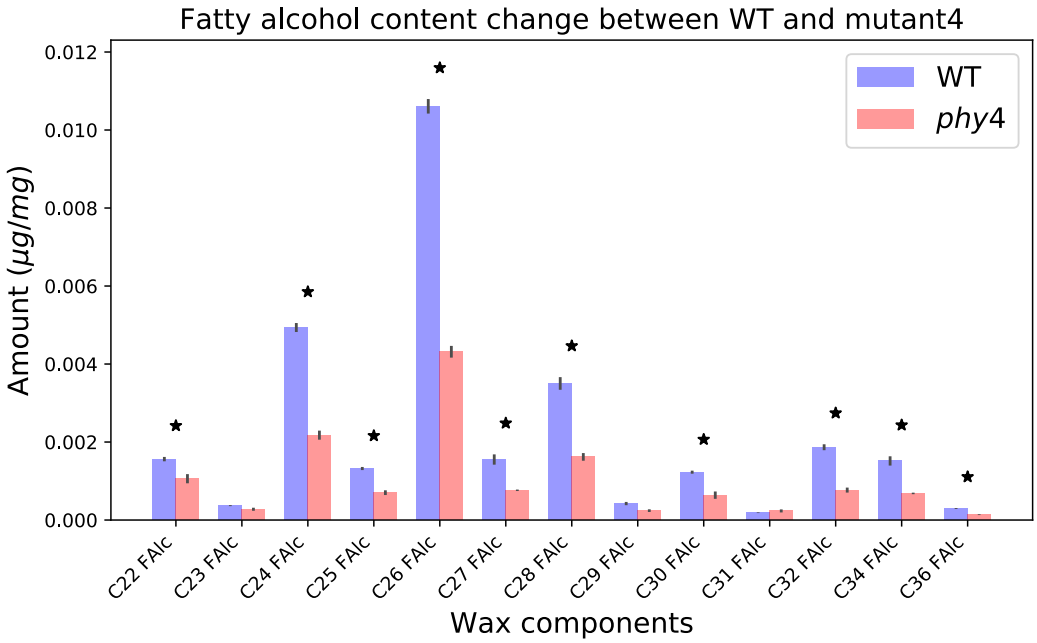

Figure 2.19. Comparisons of fatty alcohols in cuticular wax profiles in *phy4* mutants versus nonmutant plants in moss *P. patens.* Asterisks indicate significant differences (p-value < 0.05) between samples. Error bars represent standard errors.

To further investigate the underlying mechanism of the proposed relationship between light signaling and cuticle biosynthesis, we performed RNAseq on light/dark-treated leaves from wild-type and *phyb1 phyb2* mutant maize plants. In this experiment, plants were dissected to fully expose the partially-expanded maize leaf eight. The basal 2-8 cm of these leaves, which were previously light-shielded in the leaf whorl, were sampled for RNAseq analyses after three different treatment regimens: (1) immediately after dissection, and following six hour-long exposure to either (2) light or (3) dark conditions.

Three criteria were used to identify the transcripts impacting maize cuticle development that respond to PHYB signaling. These included: 1) genes differentially-expressed in wild-type plants when compared to mutants, immediately after dissection and after 6 hour light exposure; 2) genes differentially-expressed upon light exposure in wild-type plants, but not differentially-expressed in *phyb1 phyb2* double mutants; and 3) genes that showed significant interaction between the genotype and light/dark treatment in a two-way ANOVA. Only three genes remained following this strict filtering regimem (*GRMZM2G107499*, *GRMZM5G835629*, *GRMZM2G040689*). *GRMZM2G107499* is predicted to encode a gene product that is homologous to the Arabidodpsis LNK1 protein, which functions to integrate light signaling and the circadian clock. *GRMZM5G835629* is predicted to encode a protein kinase-homolog, whereas *GRMZM2G040689* is homologous to an Arabidopsis *BIFUNCTIONAL INHIBITOR*/*LTP* gene. Only the predicted maize *LTP* gene, *GRMZM2G040689*, was epidermally expressed. To investigate the function of this predicted maize LTP on cuticle biogenesis, we over-expressed this gene, tagged with the fluorescent maker *GFP,* in Arabidopsis under the *CaMV 35S* constitutive promoter. Surprisingly,

although expressed using a ubiquitous promoter, this LTP:GFP fusion protein is specifically localized to the plasma membrane/cell wall of epidermal cells (Figure 2.20A, B), which is the site of cuticle component biosynthesis and translocation. Moreover, these Arabidopsis plant over-expressing the heterologous *ZmLTP:GFP* construct show increases in the abundance of several cuticular wax components including fatty alcohols C_{26} and C_{28} (Figure 2.20C, D).

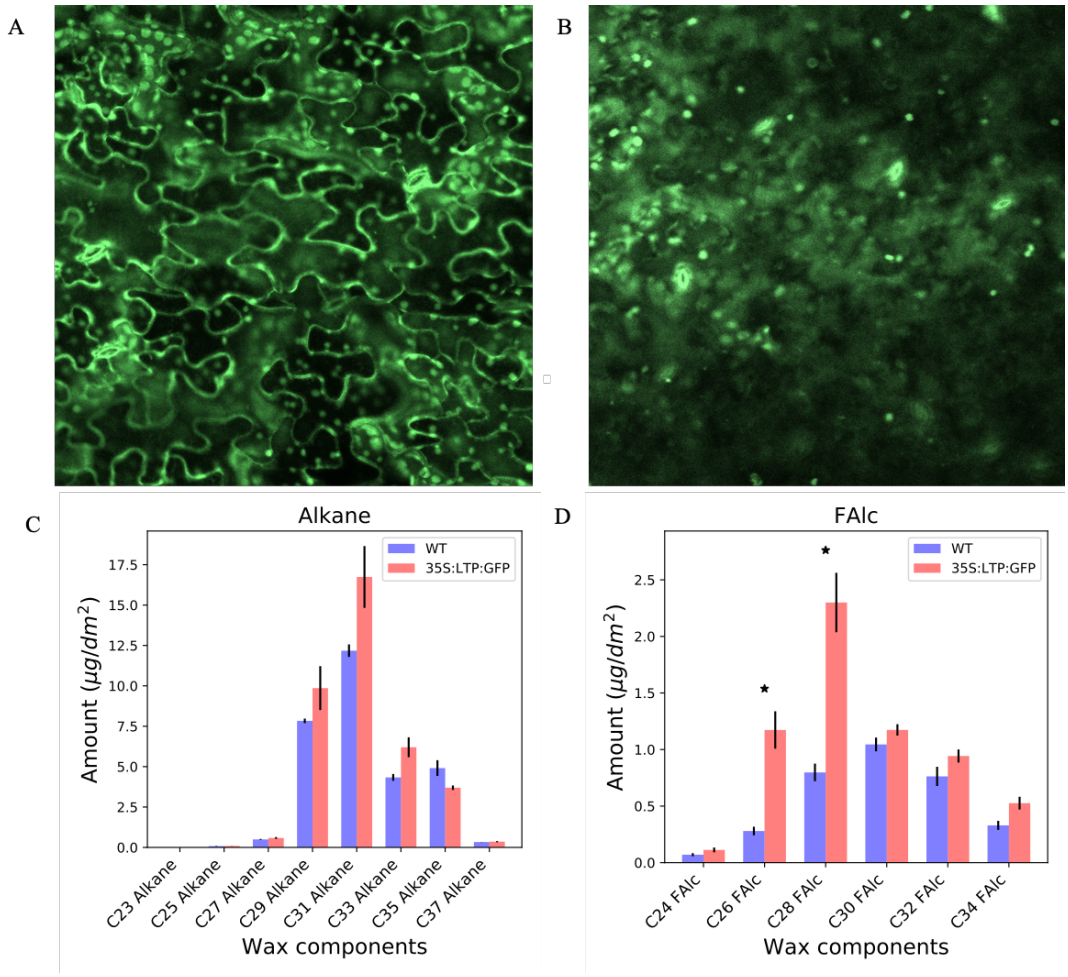

Figure 2.20. Overexpression of *GRMZM2G040689* in Arabidopsis (A) Confocal imaging reveals the accumulation of the GRMZM2G040689:GFP fusion protein in the epidermal cell plasma membrane/cell wall, which is not identified in non-transgenic wild-type plants (B). Accumulation of multiple cuticle components are altered in the *LTP* overexpression line, including alkanes (C) and fatty alcohols (D). Asterisks denote significant changes (p-value < 0.05).

PHYTOCHROMES are found in all green plants including algae (Duanmu *et al.*, 2014; Li *et al.*, 2015), although cuticles and LTPs are an evolutionary innovation of land plants (Kenrick and Crane, 1997b; Bateman *et al.*, 1998a; Edstam *et al.*, 2011; Finkina *et al.*, 2016; Salminen *et al.*, 2016; Edqvist *et al.*, 2018; Salminen *et al.*, 2018). Although light signaling-associated changes in lipid biosynthesis are present in the green algae ancestors of land plants, these light-induced lipids are not utilized to form a cuticle in algae. The evolutionary innovation of the cuticle coincided with the appearance of LTP proteins and other components of cuticle biosynthesis machinery, which are absent from algae and present in all land plants. Our data from bryophyte mosses and angiosperm grasses suggest that cuticle development in land plants is regulated by PHYTOCHROME-mediated light signaling, and that PHYTOCHROME signaling activates *LTP* expression. We propose that light-induced, PHY-mediated activation of *LTP* expression arose as a critical step enabling cuticle development during the evolution of land plants.

2.5 Materials and Methods

2.5.1 Plant material and growth conditions

B73 seeds were obtained and from the Maize Genetics Cooperation Stock Center; maize plants were grown in a 25 °C day, 20 °C night, in 60% relative humidity and 10-hour day length Percival A100 growth chambers (Percival Scientific, Perry, IA) until harvest.

Maize *phyB1 phyB2* mutants and their non-mutant siblings were obtained from J. Strable and T. Brutnell (Sheehan *et al.*, 2007) , moss *phy3* mutants and their non-mutant siblings were obtained from J. Hughes and grown as described (Whitewoods *et al.*, 2018).

2.5.2 Laser microdissection and RNAseq analysis

The unexpanded eighth leaf of the inbred B73 maize plant was harvested in three biological replicates, three plants per replicate, when the leaf was 45~55 cm, around 33 days after planting. Leaf 8 was dissected out of the whorl and segmented into 2 cm long intervals, up to 22 cm from the leaf base, seven of which (six intervals from 2 - 14 cm, and one interval from 20 – 22 cm) were fixed and paraplast–embedded for use in laser microdissection as described (Johnston *et al.*, 2014). Epidermal and internal tissues were microdissected; RNA was extracted using the PicoPure™ RNA isolation kit and linearly-amplified using the Arcturus RiboAmp ® HS PLUS RNA Amplification kit. RNA Sequencing libraries were constructed with the NEBNext Ultra™ RNA Library Prep Kit for Illumina, and the HiSEQ 2500 instrument was used for sequencing. For the RNAseq experiment on light/dark treated maize plants, 3′ RNAseq was performed. RNA was extracted with TRIzol™ Reagent and 3′ RNA sequencing library was constructed as described (Kremling *et al.*, 2018b). NextSeq 500 instrument was used for sequencing. After sequencing, reads were aligned to B73 genome RefGen_v3 with HiSAT2 (Kim *et al.*, 2015) and counted with HTSeq (Anders *et al.*, 2015).

2.5.3 Wax extraction and analysis

Waxes were extracted by submerging the tissue in pure chloroform for 60 seconds, followed by evaporation under a gentle stream of nitrogen. Dry wax samples were analyzed with GC-MS as described previously (Bourgault *et al.*, 2019a).

2.5.4 Differential expression analysis

Differential expression analysis was performed with edgeR 3.3.2 package in R (Robinson *et al.*, 2010; McCarthy *et al.*, 2012). Genes with counts fewer than 2 counts per million reads (cpm) were filtered out and analysis was carried out under FDR < 0.1 as the significant measure.

2.5.5 Weighted co-expression network analysis

The correlation between genes was performed using a modified version of Tukey's Biweight correlation (Horvath, 2011), which was later used to calculate the distance matrix. The calculations were done using WGCNA 3.3.0 package in R (Langfelder and Horvath, 2008, 2012). The distance matrix was used for the dynamic hierarchical clustering and construct the edges (connections) between nodes (genes) in the network. Network analysis of hubs and direct neighbors was done in Python 2.7 using NetworkX 1.11 module (Hagberg *et al.*, 2008).

2.5.6 Casual structure inference analysis

The epidermally upregulated genes were enriched with differential expression analysis comparing epidermal and internal samples, with Bonferroni-correction. For pair-wise comparison between adjacent epidermal sections, genes that showed significant differences in at least one comparison were input into the CSI algorithm using Cyverse (Penfold and Wild, 2011). Each pair of genes was taken as a prior, and their effects on

the section-derivative of the expression levels of all other genes gene was modeled using a Bayesian non-parametric approach (Penfold and Wild, 2011). Each gene's marginal effect on another genes expression gradient was calculated, and treated as a regulatory edge in the GRN.

2.5.7 Two-way ANOVA

The testing for significant interaction between light/dark treatment and mutant/non-mutant genotype was done with model $Y_i \sim \beta_0 + X_{1,i}\beta_1 + X_{2,i}\beta_2 + (X_{1,i} \times X_{2,i})\beta_3 + \epsilon$, where subscript *i* denotes each sample *i*. Y is the gene expression level of sample i, X_1 is the light/dark treatment, X_2 is the mutant/non-mutant genotype, and $X_1 \times X_2$ represents the interaction term between light/dark treatment and genotype. ϵ is the residual error effect assumed to be independent and identically distributed according to a normal distribution with mean of zero and variance σ^2 of one. Genes with smaller than 0.1 FDR-corrected β_3 were declared as significant in the two-way ANOVA.

2.5.8 Transgenic lines

WT Col-0 Arabidopsis plants were transformed with 35S:ZmLTP:GFP using the floral dip method (Clough and Bent, 1998). 35S:ZmLTP:GFP was made by cloning a fusion of the CMV 35S promoter driving ZmLTP into pMDC107 (Curtis and Grossniklaus, 2003). Six week old Arabidopsis were imaged on the abaxial leaf surface using LSM710 confocal microscope.

2.6 ACKNOWLEDGMENTS

The work was funded by NSF IOS award #1444507. We thank all members on the cuticle project for discussion and inputs, especially S. Matschi, M. Vasquez, A. Nguyen, M. Lin. We thank S. Leiboff for help in RNAseq processing. We also thank J. Strable, T. Brutnell for the phyB1 phyB2 double mutant and sibling maize seeds, J. Hughes for the phy3 moss mutant, and J. Cammarata for the help of growing moss.

2.7 REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Barabasi, A.L. (2013). Network science. Philos Trans A Math Phys Eng Sci **371,** 20120375.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet **5,** 101-113.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., and Stein, W.E. (1998). Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. Annual Review of Ecology and Systematics **29,** 263-292.

Beaudoin, F., Wu, X., Li, F., Haslam, R.P., Markham, J.E., Zheng, H., Napier, J.A., and Kunst, L. (2009). Functional characterization of the Arabidopsis beta-ketoacyl-coenzyme A reductase candidates of the fatty acid elongase. Plant Physiol **150,** 1174-1191.

Bird, D., Beisson, F., Brigham, A., Shin, J., Greer, S., Jetter, R., Kunst, L., Wu, X., Yephremov, A., and Samuels, L. (2007). Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J **52,** 485-498.

Bourgault, R., Matschi, S., Vasquez, M., Qiao, P., Sonntag, A., Charlebois, C., Mohammadi, M., Scanlon, M.J., Smith, L.G., and Molina, I. (2019). Changes in

lipid composition and ultrastructure associated with functional maturation of the cuticle during adult maize leaf development**,** 625343.

Chory, J., and Susek, R.E.J.C.S.H.M.A. (1994). 22 Light Signal Transduction and the Control of Seedling Development **27,** 579-614.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16,** 735-743.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133,** 462-469.

Debono, A., Yeats, T.H., Rose, J.K., Bird, D., Jetter, R., Kunst, L., and Samuels, L. (2009). Arabidopsis LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. Plant Cell **21,** 1230-1238.

Duanmu, D., Bachy, C., Sudek, S., Wong, C.H., Jimenez, V., Rockwell, N.C., Martin, S.S., Ngan, C.Y., Reistetter, E.N., van Baren, M.J., Price, D.C., Wei, C.L., Reyes-Prieto, A., Lagarias, J.C., and Worden, A.Z. (2014). Marine algae and land plants share conserved phytochrome signaling systems. Proc Natl Acad Sci U S A **111,** 15827-15832.

Edqvist, J., Blomqvist, K., Nieuwland, J., and Salminen, T.A. (2018). Plant lipid transfer proteins: are we finally closing in on the roles of these enigmatic proteins? J Lipid Res **59,** 1374-1382.

Edstam, M.M., Viitanen, L., Salminen, T.A., and Edqvist, J. (2011). Evolutionary history of the non-specific lipid transfer proteins. Mol Plant **4,** 947-964.

Fankhauser, C., and Chory, J. (1997). Light control of plant development. Annu Rev Cell Dev Biol **13,** 203-229.

Finkina, E.I., Melnikova, D.N., Bogdanov, I.V., and Ovchinnikova, T.V. (2016). Lipid Transfer Proteins As Components of the Plant Innate Immune System: Structure, Functions, and Applications. Acta Naturae **8,** 47-61.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (Los Alamos National Lab.(LANL), Los Alamos, NM (United States)).

Hooker, T.S., Millar, A.A., and Kunst, L. (2002). Significance of the expression of the CER6 condensing enzyme for cuticular wax production in Arabidopsis. Plant Physiol **129,** 1568-1580.

Horvath, S. (2011). Weighted network analysis: applications in genomics and systems biology. (Springer Science & Business Media).

Jetter, R., and Riederer, M. (2016). Localization of the Transpiration Barrier in the Epi- and Intracuticular Waxes of Eight Plant Species: Water Transport Resistances Are Associated with Fatty Acyl Rather Than Alicyclic Components. Plant Physiol **170,** 921-934.

Johnston, R., Wang, M., Sun, Q., Sylvester, A.W., Hake, S., and Scanlon, M.J. (2014). Transcriptomic analyses indicate that maize ligule development recapitulates gene expression patterns that occur during lateral organ initiation. Plant Cell **26,** 4718-4732.

Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R.J.P.m.b. (2008). The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling **67,** 547.

Kenrick, P., and Crane, P.R. (1997). The origin and early evolution of plants on land. Nature **389,** 33.

Kerstiens, G. (2006). Water transport in plant cuticles: an update. J Exp Bot **57,** 2493-2499.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods **12,** 357-360.

Kim, H., Go, Y.S., and Suh, M.C. (2018). DEWAX2 Transcription Factor Negatively Regulates Cuticular Wax Biosynthesis in Arabidopsis Leaves. Plant Cell Physiol **59,** 966-977.

Kim, H., Lee, S.B., Kim, H.J., Min, M.K., Hwang, I., and Suh, M.C. (2012). Characterization of glycosylphosphatidylinositol-anchored lipid transfer protein 2 (LTPG2) and overlapping function between LTPG/LTPG1 and LTPG2 in cuticular wax export or accumulation in Arabidopsis thaliana. Plant Cell Physiol **53,** 1391-1403.

Kolattukudy, P.E. (2001). Polyesters in higher plants. Adv Biochem Eng Biotechnol **71,** 1-49.

Kremling, K.A.G., Chen, S.Y., Su, M.H., Lepak, N.K., Romay, M.C., Swarts, K.L., Lu, F., Lorant, A., Bradbury, P.J., and Buckler, E.S. (2018). Dysregulation

of expression correlates with rare-allele burden and fitness loss in maize. Nature **555,** 520-523.

Kunst, L., and Samuels, L. (2009). Plant cuticles shine: advances in wax biosynthesis and export. Curr Opin Plant Biol **12,** 721-727.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9,** 559.

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw **46**.

Lee, S.B., Go, Y.S., Bae, H.J., Park, J.H., Cho, S.H., Cho, H.J., Lee, D.S., Park, O.K., Hwang, I., and Suh, M.C. (2009). Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules, and enhanced susceptibility to infection by the fungal pathogen Alternaria brassicicola. Plant Physiol **150,** 42-54.

Li, F.W., Melkonian, M., Rothfels, C.J., Villarreal, J.C., Stevenson, D.W., Graham, S.W., Wong, G.K., Pryer, K.M., and Mathews, S. (2015). Phytochrome diversity in green plants and the origin of canonical plant phytochromes. Nat Commun **6,** 7852.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res **40,** 4288-4297.

McFarlane, H.E., Shin, J.J., Bird, D.A., and Samuels, A.L. (2010). Arabidopsis ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. Plant Cell **22,** 3066-3075.

Mittmann, F., Brucker, G., Zeidler, M., Repp, A., Abts, T., Hartmann, E., and Hughes, J. (2004). Targeted knockout in Physcomitrella reveals direct actions of phytochrome in the cytoplasm. Proc Natl Acad Sci U S A **101,** 13939-13944.

Nawrath, C. (2002). The biopolymers cutin and suberin. Arabidopsis Book **1,** e0021.

Penfold, C.A., and Wild, D.L. (2011). How to infer gene networks from expression profiles, revisited. Interface Focus **1,** 857-870.

Pighin, J.A., Zheng, H., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L., and Samuels, A.L. (2004). Plant cuticular lipid export requires an ABC transporter. Science **306,** 702-704.

Raven, J.A., and Edwards, D. (2004). Physiological evolution of lower embryophytes: adaptations to the terrestrial environment. In The evolution of plant physiology (Elsevier), pp. 17-41.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Rowland, O., Lee, R., Franke, R., Schreiber, L., and Kunst, L. (2007). The CER3 wax biosynthetic gene from Arabidopsis thaliana is allelic to WAX2/YRE/FLP1. FEBS Lett **581,** 3538-3544.

Salminen, T.A., Blomqvist, K., and Edqvist, J. (2016). Lipid transfer proteins: classification, nomenclature, structure, and function. Planta **244,** 971-997.

Salminen, T.A., Eklund, D.M., Joly, V., Blomqvist, K., Matton, D.P., and Edqvist, J. (2018). Deciphering the Evolution and Development of the Cuticle by Studying Lipid Transfer Proteins in Mosses and Liverworts. Plants (Basel) **7**.

Samuels, L., Kunst, L., and Jetter, R. (2008). Sealing plant surfaces: cuticular wax formation by epidermal cells. Annu Rev Plant Biol **59,** 683-707.

Sheehan, M.J., Kennedy, L.M., Costich, D.E., and Brutnell, T.P. (2007). Subfunctionalization of PhyB1 and PhyB2 in the control of seedling and mature plant traits in maize. Plant J **49,** 338-353.

Sorigue, D., Legeret, B., Cuine, S., Morales, P., Mirabella, B., Guedeney, G., Li-Beisson, Y., Jetter, R., Peltier, G., and Beisson, F. (2016). Microalgae Synthesize Hydrocarbons from Long-Chain Fatty Acids via a Light-Dependent Pathway. Plant Physiol **171,** 2393-2405.

Sorigue, D., Legeret, B., Cuine, S., Blangy, S., Moulin, S., Billon, E., Richaud, P., Brugiere, S., Coute, Y., Nurizzo, D., Muller, P., Brettel, K., Pignol, D., Arnoux, P., Li-Beisson, Y., Peltier, G., and Beisson, F. (2017). An algal photoenzyme converts fatty acids to hydrocarbons. Science **357,** 903-907.

Suh, M.C., and Go, Y.S. (2014). DEWAX-mediated transcriptional repression of cuticular wax biosynthesis in Arabidopsis thaliana. Plant Signal Behav **9,** e29463.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiology **139,** 1649-1665.

Tobin, E.M., and Silverthorne, J.J.A.R.o.P.P. (1985). Light regulation of gene expression in higher plants **36,** 569-593.

Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., and Van de Peer, Y. (2009). Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. Plant Physiol **150,** 535-546.

Whitelam, G.C., and Halliday, K.J. (2007). Light and plant development. (Wiley Online Library).

Whitewoods, C.D., Cammarata, J., Nemec Venza, Z., Sang, S., Crook, A.D., Aoyama, T., Wang, X.Y., Waller, M., Kamisugi, Y., Cuming, A.C., Szovenyi, P., Nimchuk, Z.L., Roeder, A.H.K., Scanlon, M.J., and Harrison, C.J. (2018). CLAVATA Was a Genetic Novelty for the Morphological Innovation of 3D Growth in Land Plants. Curr Biol **28,** 2365-2376 e2365.

Yeats, T.H., and Rose, J.K. (2013). The formation and function of plant cuticles. Plant Physiol **163,** 5-20.

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L., and Ye, Z.H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell **20,** 2763-2782.

Chapter 3: Genomic Analyses Identify a Role for Vesicular Trafficking during Cuticular Evaporation in Maize Leaves

3.1 Abstract

Cuticles form a hydrophobic layer, covering the aboveground surface of the plant body. Plants respond to water deficit by stomatal closure, after which the majority of water loss occurs through the cuticle. Hence cuticular evaporation (CE) is a major focus toward improving plant performance under drought. To identify the underlying genetic architecture of the CE rate, we conducted a genome-wide association study (GWAS) and a transcriptome-wide association study (TWAS) on the CE rate phenotype in two environments in a panel of 321 maize inbred lines. We then used a Fisher's combined test on these GWAS and TWAS data to boost the statistical power of candidate gene nomination. Several key cuticle biosynthesis genes were associated with CE rate in maize, in addition to genes involved in multivesicular body (MVB) trafficking. Co-expression network analyses provided further evidence for the role of MVB trafficking in the regulation of CE rate. This study generated a high-dimensional dataset from a diverse population of inbred lines, and identifies target genes for reverse genetic analyses toward crop improvement under drought conditions.

3.2 Introduction

When land plants colonized the terrestrial environment, they were presented with numerous challenges including water loss in the desiccating air (Kenrick and Crane,

1997a; Raven and Edwards, 2004). Plants lose most of their water via transpiration through stomatal pores. However under drought conditions, most plants close their stomata, whereupon the hydrophobic, lipid-based, epidermal cuticle provides the major barrier against water loss (Kenrick and Crane, 1997a; Raven and Edwards, 2004; Raven *et al.*, 2005). Although a pivotal evolutionary adaptation, cuticles are not an absolute seal for water loss under stomatal closure (Kelliher *et al.*, 1995; Daszkowska-Golec and Szarejko, 2013). Thus, genetic approaches to minimize cuticular evaporation (CE) are ideal strategies toward crop improvement. In spite of this, studies of the genetic regulators of CE rate in crop plants are lacking.

Cuticles are composed of waxes and cutin (Suh *et al.*, 2005a; Kunst and Samuels, 2009b). Waxes are long-chain carbon molecules that are deposited within and on top of the cutin layer of the plant cuticle (Kolattukudy, 2001a; Nawrath, 2002a; Yeats and Rose, 2013a). Synthesis of both waxes and cutins occurs from C16/C18 lipid precursors in the endoplasmic reticulum (ER). After carbon chain elongation and further modification in the ER, these cuticle components are translocated into the apoplastic space via transporters located on the epidermal plasma membrane (Suh *et al.*, 2005a; Joubès *et al.*, 2008; Kunst and Samuels, 2009b; Yeats and Rose, 2013a). Previous studies have identified many cuticle biosynthetic enzymes, while the transportation of cuticle components is less understood. It has been hypothesized that wax components are transported in either the membranes or lumen of intracellular vesicles, from the ER through the Golgi apparatus to the trans-Golgi network, and eventually merge with the plasma membrane (Coll *et al.*, 2007; Mansbach and Siddiqi, 2010; McFarlane *et al.*, 2014). Indeed, the intracellular trafficking through the Golgi and trans-Golgi networks has been demonstrated during secretion of cuticular wax

component in plants (McFarlane *et al.*, 2014). However, additional studies are required to identify candidate genes regulating this process. The wax components and cutin monomers are later transported to the apoplastic space, through the plasma membrane, via ABCG transporters.

Genome- and transcriptome-wide association studies (i.e, GWAS and TWAS) search for significant associations that correlate the trait-of-interest with SNPs or transcripts, respectively (Yu *et al.*, 2006; Zhang *et al.*, 2010; Kremling *et al.*, 2018a). In this study, the genetic diversity within the Wisconsin Diversity Panel of maize was sampled in GWAS and TWAS analyses to identify candidate genes regulating CE rate. These combined analyses(Kremling *et al.*, 2018a) identified putative cuticle biosynthetic genes, as well as genes predicted to function in the multivesicular body (MVB) trafficking. This study provides a unique set of candidate genes, amenable to reverse genetic analyses toward the improved performance of crop plants under drought stress.

3.3 Materials and Methods

3.3.1 Experimental design

A set of 323 maize inbred lines from the Wisconsin Diversity (WiDiv) panel (Hansey *et al.*, 2011) was planted at University of California San Diego, San Diego, CA, for adult leaf cuticular evaporation (CE) rate evaluation and 3' RNA sequencing (RNAseq) tissue collection. Among the 323 lines, 89 lines were selected based on their extremely high and low CE rate in San Diego from years 2016 and 2017, in previous studies performed in UCSD (Lin *et al.* manuscript in preparation), and 234 lines were selected using the CDmean method (Rincent *et al.*, 2012) to maximize the

representation of genetic diversity from the remaining lines in the WiDiv panel after the exclusion of lines with extremely early and late flowering time (days to anthesis), in San Diego from years 2016 and 2017. The inbred lines were planted two months apart (May and July) in 2018 under standard agronomic practices. The layout of the experiment in each environment was arranged as a 17 × 19 incomplete block design. Each incomplete block contained 17 experimental lines and was augmented by the random positioning of one check (Mo17). Edge effects were reduced by planting border maize plants around the perimeter of each replicate. Experimental units were one-row plots of 3.66 m in length with 1.02 m inter-row space. At the end of each plot there was a 0.61 m alley. In each plot, 24 kernels were planted and were later thinned to approximate 12 plants.

3.3.2 Tissue collection and RNA sequencing

Plants were grown in two plot locations at different times (May and July 2018) in San Diego, CA. Leaf samples were collected from the proximal, immature and actively growing 10% - 30% of the total blade length of the unexpanded second, third or fourth fully adult leaf. One leaf sample was taken per individual plant, and three plants were sampled per plot. Tissues were frozen in liquid nitrogen right after harvesting in San Diego, CA, and shipped overnight to Ithaca, NY on dry ice.

Leaf samples from three plants were pooled and ground together, and sent to the Institute of Biotechnology at Cornell University for RNA extraction and sequencing library preparation. Samples were pooled (up to 96 multiplexity per sequencing lane), followed by sequencing on an Illumina NextSeq500 instrument. After the acquisition of raw reads, Trimmomatic (Bolger *et al.*, 2014) and cutadapt (Martin, 2011) were

used to trim off low quality reads, adaptors, random priming indices, and poly A tails. HiSAT2 (Kim *et al.*, 2015) was used to align the trimmed reads to maize genome version 4, and HTSEQ (Anders *et al.*, 2015) was used to count the reads. Samples with fewer than two million reads, or that had overall alignment rates less than 60% were discarded, and genes with fewer than one count per million in more than 10% of the samples were filtered out. Across all lines, 3984 genes had zero count per million. Pearson's correlation coefficients were calculated, using expression levels, between each pair of inbred lines within each environment. A line was removed from further analysis if the average Pearson's correlation coefficient between that line and the rest of the populations was lower than 0.7 (Hirsch *et al.*, 2014).

To confirm the identity of the 3′RNAseq samples, SNPs were called from RNAseq reads with samtools 1.9 (Li *et al.*, 2009; Li, 2011). Genotype-by-sequencing (GBS) SNP markers (Lin *et al.* manuscript in preparation) were used as cross-checks, to identify contaminant and mislabeled samples in the RNAseq data. SNPs that mapped to the same physical locations in genome version AGPv4 in the RNAseq and GBS experiments were filtered to retain biallelic SNPs (singletons and doubletons were excluded) with a call rate greater than 40%. Identity by state (IBS) was calculated between each pair of RNAseq and GBS samples, and a neighbor-joining tree was constructed based on the dissimilar matrix (*I*-IBS). RNAseq samples that were not clustered with their corresponding GBS samples were considered mislabeled or contaminated, and excluded for further analysis.

3.3.3 Phenotypic data collection

Leaf blades from the primary ear node (or one leaf above/below) were excised approximately one inch below the ligule. In plots with more than three available plants, one leaf was collected per plant, for a total of up to five leaves. Immediately after being excised, harvested leaves were collected into water to ensure submergence of the cut edges. Harvested leaves were stored in a dark room at 25 °C for two hours to ensure stomatal closure. Subsequently, leaves were wiped dry and the cut edges were sealed with parafilm. Each leaf was then hung in the dark room and weighed every 45 minutes for 180 minutes. Finally, each leaf was dried at 60 °C for four days, followed by dry weight measurement. The weight change rate in 45 minute intervals normalized by dry weight was recorded as the cuticle evaporation (CE) rate.

3.3.4 Statistical analysis

To screen the phenotypic data for significant outliers, a mixed linear model was fitted to combine CE rate from both environments. The model terms included grand mean and check as fixed effects and environment, genotype, genotype-by-environment (G×E) interaction, incomplete block within environment, and column within environment as random effects. The Studentized deleted residuals (Kutner *et al.*, 2005) generated from the mixed linear model were assessed and outliers were removed at $\alpha = 0.05$. An iterative mixed model fitting procedure was conducted for the full model with the outlier removed in ASReml-R version 3.0 (Butler *et al.*, 2009). All random terms that were not significant at $\alpha = 0.05$ were removed from the model, allowing a best-fit model to generate a best linear unbiased predictor (BLUP) for each line.

Variance component estimates obtained by fitting the full model were used to estimate heritability on a line-mean basis (Holland *et al.*, 2003; Hung *et al.*, 2012) for CE rate

across environments. Standard errors of the heritability estimates were calculated with the delta method (Lynch and Walsh, 1998; Holland *et al.*, 2003).

3.3.5 Genotypic data imputation

Genotypes from maize HapMap 3.2.1 (Bukowski *et al.*, 2017) were used to impute genotypes of selected inbred lines in the Wisconsin Diversity panel anchored by GBS SNPs. HapMap 3.2.1 contains 81,687,392 SNPs called from re-sequencing data of 1,268 maize inbred lines, and unimputed HapMap 3.2.1 genotypes with AGP_v4 coordinates were downloaded from Panzea (https://www.panzea.org/). To obtain high quality SNPs as reference, genotypes from Hapmap 3.2.1 were filtered to exclude heterozygotes and SNPs flagged as NI5 (within 5 bp of a putative indel), and retain 16,241,526 biallelic SNPs (singleton and doubleton SNPs excluded) in local LD with the GBS anchor map (Gore *et al.*, 2009; Glaubitz *et al.*, 2014) and having a SNP call rate no less than 50%. Missing SNP genotypes from HapMap 3.2.1 were imputed using Beagle 5.0 (Browning and Browning, 2016; Browning *et al.*, 2018). Imputed HapMap 3.2.1 genotypes were projected to the GBS-genotyped inbred diversity panel using Beagle 5.0. Projection was anchored by 163,557 biallelic GBS SNPs (singleton and doubleton SNPs excluded) that share physical positions and major/minor alleles with the imputed HapMap 3.2.1 SNPs. The projected genotype data were further filtered to retain SNPs with DR^2 no less than 0.8 (Browning and Browning, 2016) and minor allele counts of at least 40 (minor allele frequency 6.23%). The final complete set contains 10,860,644 high-quality SNP markers, of which 3,459,038 were retained after linkage disequilibrium (LD) pruning ($r^2 \leq 0.9$) of the complete marker data set in PLINK version 1.09_beta5 (Purcell *et al.*, 2007).

3.3.6 Genome-wide association study

Utilizing the best linear unbiased predictors (BLUPs) for CE rate in two environments collected in 2018, a GWAS analysis was conducted using 3,459,038 high-quality SNPs imputed from the HapMap 3.2.1 data with R package GAPIT (Zhang *et al.*, 2010; Lipka *et al.*, 2012). No principal components were selected, evaluated by Bayesian information criterion. A kinship matrix calculated with the IBS method(Bradbury *et al.*, 2007) was included in the model to adjust for population structure and family relatedness. P-values were adjusted as false discovery rate (FDR) for multiple testing correction, and SNPs with lower than 0.1 FDR were claimed as significant.

3.3.7 Transcriptome-wide association study

Aligned and counted RNAseq reads were first log transformed with base two. The reads were then fitted with mixed linear models with grand mean and check as fixed effects, block and column within environment, sequencing plate/lane, environment, genotype and G×E as random effects. The output (BLUPs) were then input into R package PEER (Stegle *et al.*, 2010; Stegle *et al.*, 2012) to remove unobserved latent variables that skew gene expression levels. PEER residuals were output as representations of unbiased "real" gene expression levels to use in the transcriptome-wide association study (TWAS). Outlier lines were removed at $\alpha = 0.05$ for Studentized deleted residuals (Kutner *et al.*, 2005) in the null model with only grand mean. Two principal components (PCs) were selected in the null model by Bayesian information criterion to control for population structure. An F-test between the null model with two PCs and the full model with two PCs plus the transcript level was

conducted for each gene. The P-value for each gene was recorded and adjusted for multiple testing by false discovery rate (FDR) as described in the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

3.3.9 Fisher's combined tests

To perform Fisher's combined test, 346,341 most associated SNPs (top 10%) in GWAS were assigned to nearest genes, and P-values of each of the top ranking SNPs in GWAS were combined with P-values of their nearest genes in TWAS. For genes with unobserved expression levels, and thus not tested in TWAS, P-values were combined with GWAS P-values for the Fisher's combined test. Fisher's combined tests were implemented using the paired P-values in the sumlog method in the metap R package (Dewey, 2016). The top 1% most significant genes in the Fisher's combined test were considered as candidate genes regulating CE rate in the maize leaf.

3.4 Results and Discussion

3.4.1 Phenotypic variability of cuticular evaporation rate

Cuticular evaporation (CE) describes the water loss through the leaf epidermis upon stomatal closure, which is assumed to involve evaporation through the cuticle. CE rate is therefore defined as the water loss per unit time and unit dry weight of the leaf. In this study, CE rates were measured in 323 inbred lines sampled from the Wisconsin Diversity Panel grown in two environments in San Diego, CA, and best unbiased linear predictions (BLUPs) for this trait were calculated as described in Materials and Methods. The CE rate BLUPs show a wide range of distribution, with the majority normally distributed (Figure 3.1). CE rate showed a heritability on a line-mean basis

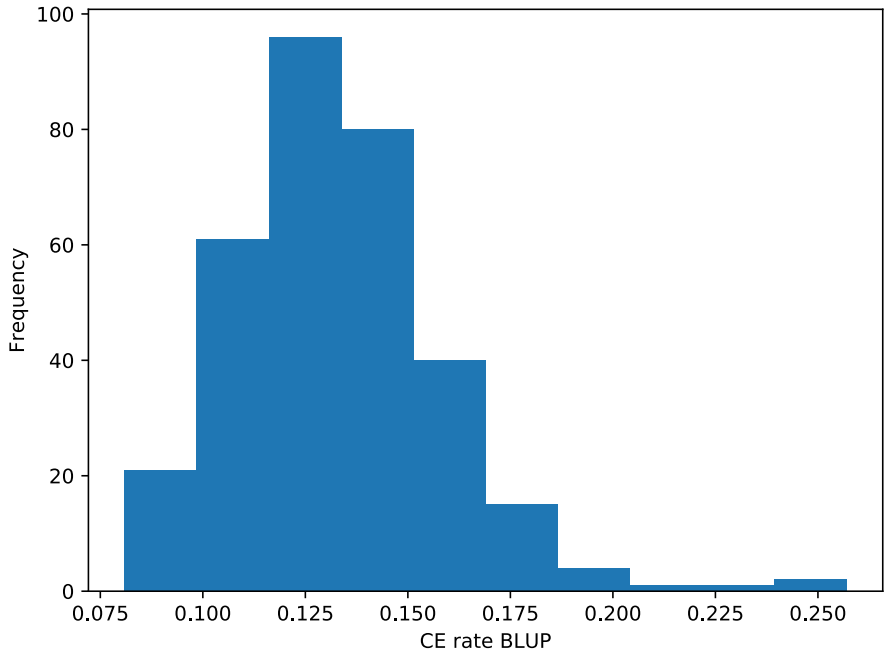

Figure 3.1. Histogram of CE rate BLUPs. The majority of CE rate BLUPs is normally distributed, with lines with CE rate BLUPs higher than 0.2 as the right tail.

($\widehat{h_l^2}$) of 0.73, indicating that this trait is strongly influenced by genetics and is amenable to GWAS and TWAS studies.

3.4.2 Genome-wide association study identifies MVB trafficking genes as potential CE rate regulators

A GWAS was performed to analyze the genetic architecture of CE rate using a total of 3,459,038 single-nucleotide-polymorphisms (SNPs) mapped and imputed within the population as genetic markers. Given the relatively small sample size (n = 321) used in the GWAS analysis, a genome-wide false discovery rate (FDR) cutoff of 0.1 was used to select for candidate SNPs, to ensure the detection of weaker effects. Results of the GWAS analysis is shown in Figure 3.2.

A total of seven SNPs passed the 0.1 FDR cutoff, distributed on chromosomes 4 and 10 (Figure 3.2). The five significant SNPs on chromosome 10 physically mark a window of 551.4 kb comprising eleven predicted maize genes. The other two significant SNPs on chromosome 4 span a 251.4 kb window, containing five genes (Table 3.1). Among these sixteen genes, interesting candidates include *Zm00001d049479*, whose Arabidopsis homolog *VACUOLAR PROTEIN SORTING4* (*VPS4*) encoding a protein involved in the MVB pathway (Reyes *et al.*, 2014). Cuticle components are hydrophobic lipids that translocate from the ER to the extracellular space. It has been proposed that this process may comprise two parts: 1) vesicular transport from ER to the plasma membrane; 2) from the plasma membrane to the apoplastic space. A previous study demonstrated that wax secretion requires Golgi- and trans-Golgi network-mediated vesicle trafficking (McFarlane *et al.*, 2014), providing support for the hypothesis that wax components are delivered from the ER

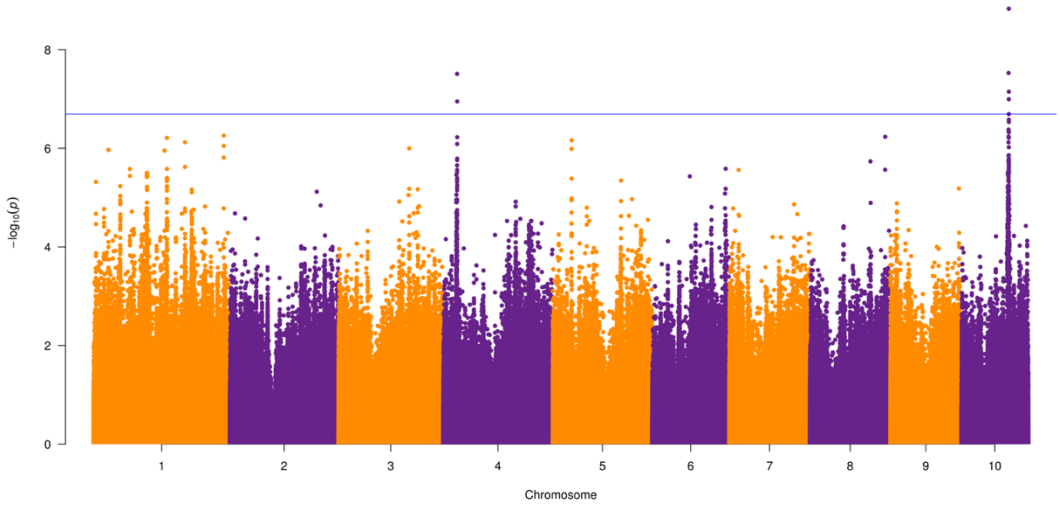

Figure 3.2. Manhattan Plot of GWAS for CE rate in maize. Each point represent a SNP, and the x axis shows the chromosomal position of the corresponding SNP. The y axis shows the negative $\log_{10}$ transformed P-values. The blue horizontal line is the -$\log_{10}$ p-value for SNPs that passed the 0.1 FDR cutoff. Seven SNPs are significantly associated with CE rate.

Table 3.1. Gene in the window of significant GWAS SNPs

Maize gene ID	Chromosome	Maize gene annotation	Arabidopsis homolog annotation
Zm00001d025149	10	*KINESIN-LIKE PROTEIN KIN-12A*	*PHRAGMOPLAST-ASSOCIATED KINESIN-RELATED PROTEIN, PUTATIVE*
Zm00001d025151	10	*NA*	*NA*
Zm00001d025152	10	*PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN*	*NA*
Zm00001d025153	10	*PHOSPHOLIPID-TRANSPORTING ATPASE 2*	*AMINOPHOSPHOLIPID ATPASE 2*
Zm00001d025154	10	*NA*	*NA*
Zm00001d025155	10	*RETROVIRUS-RELATED POL POLYPROTEIN LINE-1*	*NA*
Zm00001d025156	10	*PHOSPHOLIPID-TRANSPORTING ATPASE 2*	*AMINOPHOSPHOLIPID ATPASE 2*
Zm00001d025160	10	*60S RIBOSOMAL PROTEIN L35*	*RIBOSOMAL L29 FAMILY PROTEIN*

Table 3.1 continued. Gene in the window of significant GWAS SNPs

Maize gene ID	Chromosome	Maize gene annotation	Arabidopsis homolog annotation
Zm00001d025161	10	*PUTATIVE ARM REPEAT-CONTAINING PROTEIN CONTAINING FAMILY PROTEIN*	*ARM REPEAT SUPERFAMILY PROTEIN*
Zm00001d025162	10	*NA*	
Zm00001d049474	4	*PROTEIN VAC14 HOMOLOG*	*NA*
Zm00001d049475	4	*DOUBLE-STRAND BREAK REPAIR PROTEIN MRE11*	*DNA REPAIR AND MEIOSIS PROTEIN (MRE11)*
Zm00001d049476	4	*Z1A ALPHA ZEIN PROTEIN*	*NA*
Zm00001d049477	4	*NA*	*NA*
Zm00001d049479	*4*	*REGULATOR OF VPS4 ACTIVITY IN THE MVB PATHWAY PROTEIN*	*REGULATOR OF VPS4 ACTIVITY IN THE MVB PATHWAY PROTEIN*

to the Golgi apparatus, and further transported to the plasma membrane. Hence the maize *VPS4* homolog *Zm00001d049479* may be involved in the cuticle component trafficking pathway, regulating CE rate in maize. Additionally, *Zm00001d025149* on chromosome 10 has Arabidopsis homolog involved in the function of KINESIN, which may act as motor proteins for the transportation of MVBs.

3.4.3 Transcriptome-wide association study suggests that the CE rate is transcriptionally regulated

While GWAS tests for significant association between a trait and a SNP marker, TWAS aims to identify the association between a trait and the expression level of a particular gene transcript (Kosma *et al.*, 2015). Thus, TWAS can complement GWAS, especially when the transcriptomic data are specifically derived from functionally-relevant organs/tissues/cells. This TWAS analyses sampled RNA from sections of immature leaf blade, where cuticle biogenesis genes are assumed to be expressed (Qiao *et al.*, Chapter 2 of this dissertation). After RNA sequencing, the association between gene expression levels and CE rate was investigated; results of this TWAS for CE rate are summarized in Figure 3.3.

Two maize transcripts (i.e. gene IDs *Zm00001d005087* and *Zm00001d042888*) were significantly associated (0.1 FDR) with CE rate in our TWAS. *Zm00001d005087* encodes a G2-LIKE TRANSCRIPTION FACTOR 35 (GLK35), a member of the homeodomain-like superfamily. The function of GLK35 has not been previously studied, although it is a predicted transcription factor. The second transcript identified in the TWAS is encoded by gene *Zm00001d042888*, which is unannotated in maize.

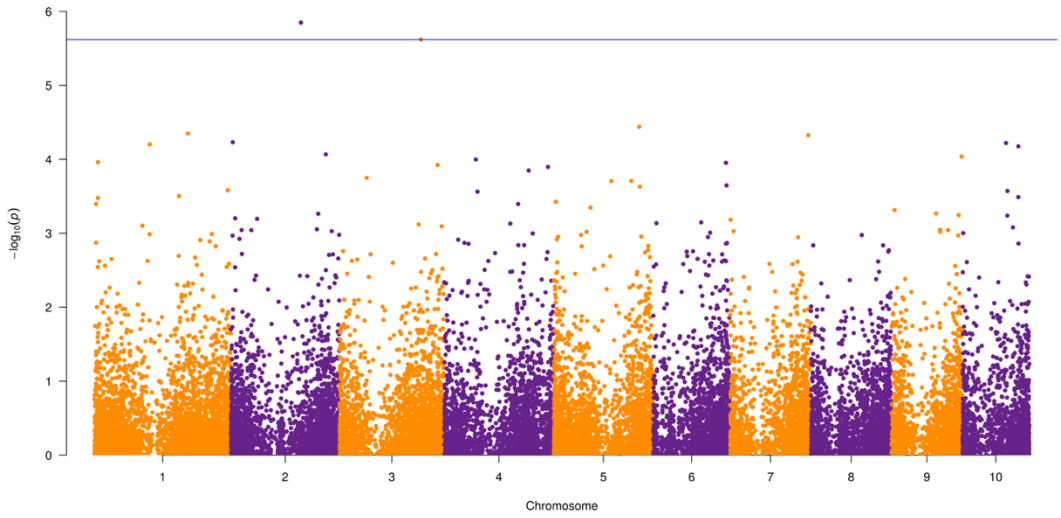

Figure 3.3. Manhattan plot for TWAS for CE rate in maize. Each point represents a transcript, and the x axis shows the chromosomal position of the corresponding gene. The y axis shows the negative $\log_{10}$ transformed P-values. The blue horizontal line is the -log10 P-value for the transcript that passed the 0.1 FDR cutoff. Two gene transcripts are significantly associated with CE rate.

However, the Arabidopsis homolog of *Zm00001d042888* encodes RIBOSOMAL RNA PROCESSING 4, a subunit of RNA exosome that is required for the degradation of abnormal transcripts (Reyes *et al.*, 2014). Thus the GWAS and TWAS identified unique sets of candidate loci associated with CE rate in maize. The candidates identified in the TWAS may be regulating the transcription of the GWAS loci identified, or vice versa.

3.4.4 Fisher's combined probability test selects candidates implicated in cuticle biosynthesis and MVB trafficking

To increase the statistical power of candidate gene detection, GWAS and TWAS results can be fused using a Fisher's combined probability test (Kremling *et al.*, 2018a). In this process, the P-values of the most associated SNPs (top 10%) in GWAS were assigned to nearest genes, and P-values of each of the top ranking SNPs in GWAS were combined with P-values of their nearest genes in TWAS. Subsequently, these two P-values were combined to test if a gene is associated with CE rate. Utilizing two P-values per SNP, one derived from GWAS and one from TWAS, can thus synergize the association of a SNP to CE rates. These results of our Fisher's combined test are summarized in Figure 3.4. The top 1% most-significant SNPs were investigated to search for candidate genes regulating CE rate, and the potential candidates can be categorized into two major classes.

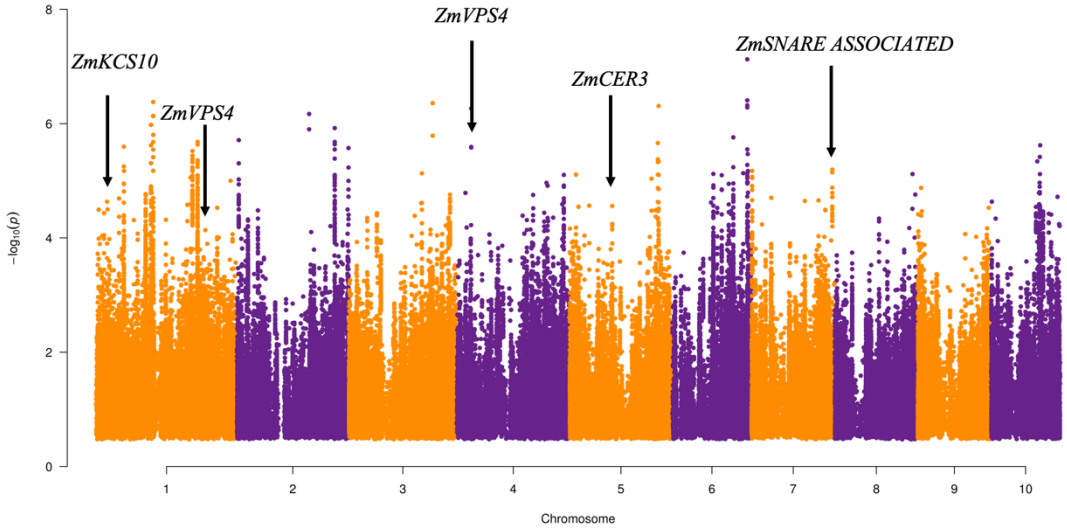

Figure 3.4. Manhattan plot of a Fisher's combined test of GWAS and TWAS for CE rate. Each point represent a SNP, and the x axis shows the chromosomal position of the corresponding SNP. The y axis shows the negative $\log_{10}$ transformed P-values.

Three maize genes *Zm00001d015477* (*ZmECERIFERUM3*, or *ZmCER3*), *Zm00001d027904* (*Zm3-KETOACYL-COA-SYNTHASE10*, or *ZmKCS10*), and *Zm00001d032284* (ZmHOTHEAD), have homologs implicated in cuticle biogenesis in Arabidopsis (Kurdyukov *et al.*, 2006; Wu *et al.*, 2011; Bernard *et al.*, 2012). As the main barrier to water loss through the epidermis, the cuticle is critical to CE rate. However, to date the detailed cuticle components that contribute to the prevention of water loss are not identified. The gene candidate identified, ZmKCS10, have Arabidopsis homologs involved in the fatty acid elongation process (Wu *et al.*, 2011; Bernard *et al.*, 2012), which suggests a role for long-chain cuticle components in the regulation of CE rate. On the other hand, ZmCER3 is involved in alkane synthesis, suggesting a role of alkanes in CE rate regulation. Additionally, HOTHEAD is required for formation of cutin monomers (Kurdyukov *et al.*, 2006). Cutin is a polymer matrix that interlocks with the outer cell wall of epidermal cells, thereby providing an interface for wax deposition. As such, cutin has long been hypothesized to perform a strategic role in water loss prevention. The association between CE rate and HOTHEAD provides additional support for the role of cutin in water loss prevention.

Three additional candidate genes (*Zm00001d022250*, *Zm00001d022364*, and *Zm00001d049479*) identified in the Fisher's combined test have homologs involved in endomembrane vesicle trafficking. Both Zm00001d022250 and *Zm00001d022364* have Arabidopsis homolog encoding SNARE associated Golgi proteins. SNARE proteins dock vesicles to membranes before they fuse (Chen and Scheller, 2001), whereas *Zm00001d049479* encodes the VPS4 homolog also identified in our GWAS (Figure 3.2). Hence we propose that MVB trafficking plays important roles in

transporting cuticle components, which in turn influences non-stomatal CE rate. Moreover, MVB trafficking requires cytoskeletal components (Hehnly and Stamnes, 2007; Gurel *et al.*, 2014). Two additional genes (*Zm00001d051688* and *Zm00001d025117*) identified in the Fisher's combined test are annotated as *ZmFIMBRIN HOMOLOG1* and *ZmACTIN ASSOCIATED PROTEIN*. In Arabidopsis, FIMBRINs function during actin bundling. Taken together, these data implicate a role for the cytoskeletal component actin in the regulation of CE rate in maize.

3.4.5 Camoco-aided gene discovery

To further facilitate the discovery of genes regulating maize CE rate, gene co-expression networks were constructed using the transcriptomic data generated in this study. GWAS can identify significant SNPs, however, the identification of candidate genes near significant SNPs can be challenging. Camoco is a computational tool that integrates GWAS-identified SNPs with functional information (i.e. Gene Ontology or GO) derived from gene expression data, to facilitate the nomination of gene candidates linked to a significant SNP (Schaefer *et al.*, 2018). Remarkably, Camoco also identified ZmVPS4 as a candidate gene, further supporting the hypothesis that MVB trafficking impacts CE rate in maize.

In summary, phenotypic analysis uncovered a wide range of variation for CE rates in 321 maize inbred lines. Exploiting this natural variation in an important agronomic trait, GWAS and TWAS identified several interesting candidate genes regulating CE rate. Subsequently, a Fisher's combined test was used to synergize these data and demonstrated a significant association between CE rate and MVB trafficking. Statistical correlation of co-expression analyses with GO information likewise

supported a role for MVB trafficking in CE rate regulation. This provides insight into the process whereby cuticle components are exported from plant epidermal cells, and identifies candidate genes for reverse genetic analyses of drought tolerance in maize.

3.5 REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Benjamini, Y., and Hochberg, Y.J.J.o.t.R.s.s.s.B. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing **57,** 289-300.

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.D., Haslam, R.P., Napier, J.A., Lessire, R., and Joubes, J. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that Arabidopsis ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. Plant Cell **24,** 3106-3118.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30,** 2114-2120.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S.J.B. (2007). TASSEL: software for association mapping of complex traits in diverse samples **23,** 2633-2635.

Browning, B.L., and Browning, S.R.J.T.A.J.o.H.G. (2016). Genotype imputation with millions of reference samples **98,** 116-126.

Browning, B.L., Zhou, Y., and Browning, S.R.J.T.A.J.o.H.G. (2018). A one-penny imputed genome from next-generation reference panels **103,** 338-348.

Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., Wang, B., Xu, D., Yang, B., and Xie, C.J.G. (2017). Construction of the third-generation Zea mays haplotype map **7,** gix134.

Butler, D., Cullis, B.R., Gilmour, A., Gogel, B.J.T.S.o.Q., Department of Primary Industries, and Fisheries, B. (2009). ASReml-R reference manual.

Chen, Y.A., and Scheller, R.H.J.N.r.M.c.b. (2001). SNARE-mediated membrane fusion **2,** 98.

Coll, E.P., Kandt, C., Bird, D.A., Samuels, A.L., and Tieleman, D.P.J.T.J.o.P.C.B. (2007). The distribution and conformation of very long-chain plant wax components in a lipid bilayer **111,** 8702-8704.

Daszkowska-Golec, A., and Szarejko, I. (2013). Open or close the gate - stomata action under the control of phytohormones in drought stress conditions. Front Plant Sci **4,** 138.

Dewey, M. (2016). metap: Meta-analysis of significance values. R package version 0.7.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S.J.P.o. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline **9,** e90346.

Gore, M.A., Chia, J.-M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills, G.S., and Ross-Ibarra, J.J.S. (2009). A first-generation haplotype map of maize **326,** 1115-1117.

Gurel, P.S., Hatch, A.L., and Higgs, H.N. (2014). Connecting the cytoskeleton to the endoplasmic reticulum and Golgi. Curr Biol **24,** R660-R672.

Hansey, C.N., Johnson, J.M., Sekhon, R.S., Kaeppler, S.M., and Leon, N.d.J.C.S. (2011). Genetic diversity of a maize association population with restricted phenology **51,** 704-715.

Hehnly, H., and Stamnes, M. (2007). Regulating cytoskeleton-based vesicle motility. FEBS Lett **581,** 2112-2118.

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2014). Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26,** 121-135.

Holland, J.B., Nyquist, W.E., and Cervantes-Martínez, C.T.J.P.b.r. (2003). Estimating and interpreting heritability for plant breeding: an update **22,** 9-112.

Hung, H., Browne, C., Guill, K., Coles, N., Eller, M., Garcia, A., Lepak, N., Melia-Hancock, S., Oropeza-Rosas, M., and Salvo, S.J.H. (2012). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population **108,** 490.

Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R.J.P.m.b. (2008). The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling **67,** 547.

Kelliher, F., Leuning, R., Raupach, M., Schulze, E.-D.J.A., and Meteorology, F. (1995). Maximum conductances for evaporation from global vegetation types **73,** 1-16.

Kenrick, P., and Crane, P.R.J.N. (1997). The origin and early evolution of plants on land **389,** 33.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods **12,** 357-360.

Kolattukudy, P.E. (2001). Polyesters in higher plants. In Biopolyesters (Springer), pp. 1-49.

Kosma, D.K., Rice, A., and Pollard, M. (2015). Analysis of aliphatic waxes associated with root periderm or exodermis from eleven plant species. Phytochemistry **117,** 351-362.

Kremling, K., Diepenbrock, C., Gore, M., Buckler, E., and Bandillo, N.J.b. (2018). Transcriptome-wide association supplements genome-wide association in Zea mays**,** 363242.

Kunst, L., and Samuels, L.J.C.o.i.p.b. (2009). Plant cuticles shine: advances in wax biosynthesis and export **12,** 721-727.

Kurdyukov, S., Faust, A., Trenkamp, S., Bar, S., Franke, R., Efremova, N., Tietjen, K., Schreiber, L., Saedler, H., and Yephremov, A. (2006). Genetic and biochemical evidence for involvement of HOTHEAD in the biosynthesis of long-chain alpha-,omega-dicarboxylic fatty acids and formation of extracellular matrix. Planta **224,** 315-329.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). Applied linear statistical models. (McGraw-Hill Irwin Boston).

Lewontin, R.C. (1988). On measures of gametic disequilibrium. Genetics **120,** 849-852.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27,** 2987-2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25,** 2078-2079.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics **28,** 2397-2399.

Lynch, M., and Walsh, B. (1998). Genetics and analysis of quantitative traits. (Sinauer Sunderland, MA).

Mansbach, C.M., and Siddiqi, S.A.J.A.r.o.p. (2010). The biogenesis of chylomicrons **72,** 315-333.

Martin, M.J.E.j. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads **17,** 10-12.

McFarlane, H.E., Watanabe, Y., Yang, W., Huang, Y., Ohlrogge, J., and Samuels, A.L.J.P.p. (2014). Golgi-and trans-Golgi network-mediated vesicle trafficking is required for wax secretion from epidermal cells **164,** 1250-1260.

Nawrath, C. (2002). The biopolymers cutin and suberin. Arabidopsis Book **1,** e0021.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J.J.T.A.j.o.h.g. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses **81,** 559-575.

Raven, J.A., and Edwards, D. (2004). Physiological evolution of lower embryophytes: adaptations to the terrestrial environment. In The evolution of plant physiology (Elsevier), pp. 17-41.

Raven, P.H., Evert, R.F., and Eichhorn, S.E. (2005). Biology of plants. (Macmillan).

Reyes, F.C., Buono, R.A., Roschzttardtz, H., Di Rubbo, S., Yeun, L.H., Russinova, E., and Otegui, M.S. (2014). A novel endosomal sorting complex required for transport (ESCRT) component in Arabidopsis thaliana controls cell expansion and development. J Biol Chem **289,** 4980-4988.

Rincent, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., Rodriguez, V.M., Moreno-Gonzalez, J., Melchinger, A., and Bauer, E.J.G. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (Zea mays L.) **192,** 715-728.

Schaefer, R.J., Michno, J.M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., and Myers, C.L. (2018). Integrating Coexpression Networks with GWAS to Prioritize Causal Genes in Maize. Plant Cell **30,** 2922-2942.

Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol **6,** e1000770.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc **7,** 500-507.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiol **139,** 1649-1665.

Wu, R., Li, S., He, S., Wassmann, F., Yu, C., Qin, G., Schreiber, L., Qu, L.J., and Gu, H. (2011). CFL1, a WW domain protein, regulates cuticle development by modulating the function of HDG1, a class IV homeodomain transcription factor, in rice and Arabidopsis. Plant Cell **23,** 3392-3411.

Yeats, T.H., and Rose, J.K.J.P.p. (2013). The formation and function of plant cuticles **163,** 5-20.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet **38,** 203-208.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. (2010).

Mixed linear model approach adapted for genome-wide association studies. Nat Genet **42,** 355-360.

Chapter 4: Machine Learning Enables High-Throughput Phenotyping for Analyses of the Genetic Architecture of Bulliform Cell Patterning in Maize

4.1 Abstract

Bulliform cells comprise specialized cell types that develop on the adaxial (upper) surface of grass leaves, and are patterned to form linear rows along the proximodistal axis of the adult leaf blade. Bulliform cell patterning affects leaf angle and is presumed to function during leaf rolling, thereby reducing water loss during temperature extremes and drought. In this study, epidermal leaf impressions were collected from a genetically and anatomically diverse population of maize inbred lines. Subsequently, convolutional neural networks were employed to measure microscopic, bulliform cell-patterning phenotypes in high-throughput. A genome-wide association study, combined with RNAseq analyses of the bulliform cell ontogenic zone, identified candidate regulatory genes affecting bulliform cell column number and cell width. This study is the first to combine machine learning approaches, transcriptomics, and genomics to study bulliform cell patterning, and the first to utilize natural variation to investigate the genetic architecture of this microscopic trait. In addition, this study provides insight toward the improvement of macroscopic traits such as drought resistance and plant architecture in an agronomically important crop plant.

4.2 Introduction

Drought stress remains a serious challenge to agronomic production (Ort and Long, 2014); land plants have evolved multiple mechanisms for water conservation since their invasion of the terrestrial environment more than 450 million years ago (Kenrick and Crane, 1997a; Raven and Edwards, 2004). Grasses are staple crops for human subsistence and have evolved specific epidermal cells types (i.e. bulliform cells) to reduce water loss during heat and drought (Price *et al.* 1997; Kadioglu and Terzi 2007; Hu *et al.* 2010; Hsiao *et al.* 1984). Bulliform cells are enlarged parenchymatous structures arranged in tandem clusters that form linear columns along the proximodistal leaf axis (Becraft *et al.*, 2002a; Bennetzen and Hake, 2008). During heat and/or water stress, bulliform cells are proposed to shrink dramatically in size along the adaxial (top) leaf surface. This asymmetric decrease in leaf surface area is a proposed mechanism for leaf rolling, consequently reducing water loss from the leaf epidermis (Hsiao *et al.*, 1984; Price *et al.*, 1997; Dai *et al.*, 2007; Kadioglu and Terzi, 2007; Hu *et al.*, 2010). Bulliform cell number and density are also correlated with leaf angle, thus impacting plant architecture. Rice bulliform cell patterning mutants such as Rice outermost cell-specific gene5 (Roc5) over-produce bulliform cells, have more upright leaves, which is a desirable agronomic trait enabling dense planting (Zou *et al.*, 2011).

Despite the inherent interest in bulliform cell patterning to both plant developmental biologists and breeders, previous studies have focused on either the cell-specific transcriptomes or reverse genetics analyses of mature-staged bulliform cells. For example, a study in rice showed that bulliform cells express around 16,000 genes, far more than the median of 8,831 genes identified in RNAseq analyses of over 40 distinct cell types (Jiao *et al.*, 2009). Coincidentally, reverse genetic studies reveal that

mutations in genes implicated in a diverse array of biological processes can condition bulliform cell phenotypes. For example, the phytohormones brassinosteroid, gibberellin, and auxin both function during bulliform cell patterning in rice (Dai *et al.*, 2007; Fujino *et al.*, 2008; Chen *et al.*, 2015b), whereas some leaf-rolling mutants have supernumerary bulliform cells and others develop ectopic bulliform cells on the abaxial (bottom) side of the leaf (Itoh *et al.*, 2008a; Hibara *et al.*, 2009a; Zhang *et al.*, 2009b; Li *et al.*, 2010a). Aside from defects in adaxial/abaxial patterning, some leaf rolling mutants are also impaired in water transport (Fang *et al.*, 2012a), or in the production of a vacuolar ATPase (Xiang *et al.*, 2012). Despite these genetic analyses of bulliform development, no studies have been performed on the natural variation of bulliform cell patterning in a staple crop plant such as maize.

Elucidating the genetic architecture controlling natural variation of maize bulliform cell patterning is fraught with challenges. Although bulliform cells influence a wide range of macroscopic traits such as leaf rolling and leaf angle, bulliform cell patterning is a microscopic phenotype. Historically, epidermal cells are typically analyzed by scanning electron microscopy (SEM) (Becraft *et al.*, 2002a) or light-imaging of epidermal glue-impressions (Bennetzen and Hake, 2008). Although SEM is not amenable to high-throughput phenotyping of large plant populations, epidermal glue-impressions are relatively easy to generate in high volume and can be stored for extended periods, thereby preserving cellular structures in great detail (Bennetzen and Hake, 2008).

Another bottleneck to high-throughput phenotyping of microscopic epidermal traits is the quantification of cell profiles after image acquisition. Machine learning strategies

such as convolutional neural networks (CNNs) are widely used for image processing; advances in modern technology have enabled the optimization of complex machine learning models comprising millions of parameters (LeCun *et al.*, 1995; LeCun *et al.*, 1998; Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2014; Zeiler and Fergus, 2014; Szegedy *et al.*, 2015; He *et al.*, 2016). Semantic segmentation of microscopic images via CNNs can significantly decrease the labor and time required to manually score such phenotypes in large-scale genetic studies. Special CNN algorithms such as U-net enable the efficient use of context information of image pixels, thereby reducing the otherwise daunting workload of manually tracing cell anatomical patterns into a matter of seconds (Ronneberger *et al.*, 2015).

In this study, leaf epidermal glue-impressions were collected from a genetically diverse panel of nearly 500 maize inbred lines, and U-nets were utilized to quantify bulliform cell patterning phenotypes from over 60,000 leaf images within this population. A genome-wide association study (GWAS) (Yu *et al.*, 2006; Lipka *et al.*, 2012) was then performed to identify loci associated with bulliform cell column number and width. In addition, the ontogeny of bulliform cell development in the expanding maize leaf was analyzed, which informed the stage-specific isolation of mRNA from the region of bulliform cell initiation and differentiation in the developing maize leaf. Considering both these GWAS and transcriptomic data, we propose candidate genes responsible for bulliform cell patterning in maize.

4.3 Materials and Methods

4.3.1 Bulliform cell ontogeny and RNA sequencing

Seeds of maize inbred line B73 (accession number: PI 550473) were obtained from the Maize Genetics Cooperation Stock Center. Three replicates of B73 plants were grown in Percival A100 growth chambers with 10-hour day length at temperatures 25 °C day, 20 °C night, and relative humidity of 60%. Plants were grown for 33 days, when the partially elongated leaf eight was 50-55 cm long. Leaf eight was dissected out of the whorl and EXAFLEX® Vinyl Polysiloxane Impression Material (Injectable) was applied onto the basal 5 cm of the blade to make epidermal glue-impressions.

Total RNA was isolated from the 0 – 2 mm region distal to the ligule of the expanding leaf 8 using the TRIzol™ Reagent in three replicates. The NEBNext UltraTM RNA Library Prep Kit for Illumina was used to construct sequencing libraries. The Illumina HiSEQ 2500 instrument was used for 150 bp paired-end sequencing. After sequencing, reads were aligned to B73 version 4 genome with HiSAT2 (Kim *et al.*, 2015) and counted with HTSeq (Anders *et al.*, 2015).

4.3.2 Differential gene expression analysis

Differential gene expression analysis was performed in R with the edgeR 3.3.2 package (Robinson *et al.*, 2010; McCarthy *et al.*, 2012) comparing the transcriptomes of the 2 mm and 15 - 35 mm regions distal to the ligule. Gene expression levels were normalized against library sizes. The default generalized linear model was used to call differential expressions. Genes with false discovery rate (FDR) less than 0.10 were declared as being significantly differentially expressed.

4.3.3 Experimental design

A set of 468 maize inbred lines sampled from the Wisconsin Diversity (WiDiv) panel (Hirsch *et al.*, 2014) (Table S1) were evaluated for bulliform cell patterning traits in adult leaves. The inbred lines were planted at the Maricopa Agricultural Center, Maricopa, AZ, and the University of California San Diego, San Diego, CA in 2017. The layout of the experiment in each location was arranged as a 18 x 26 incomplete block design. Each incomplete block of 18 experimental lines was augmented by the random positioning of two check inbred lines (N28HT and Mo17). The entire experiment of 468 unique inbred lines plus checks was grown as a single replicate in each location. Edge effects were reduced by planting border maize plants around the perimeter of each replicate. Experimental units were one-row plots of 3.05 m (Maricopa) and 4.88 m (San Diego) in length with 1.02 m inter-row space. At the end of each plot there was a 0.91 m alley. Twenty-four kernels were planted in each plot, which were later thinned to approximately 12 plants per plot.

4.3.4 Leaf epidermal phenotypic data collection

Plants were grown in two environments under standard agronomic practices during the summer of 2017: San Diego, CA and Maricopa, AZ. Leaf samples were taken from five plants per inbred line (plot) when at least half of the plants in that plot were at anthesis. Each leaf sample was taken midway between the ligule and the tip of the blade of the primary ear node leaf, or from one leaf younger. Midrib and margins were removed from the leaf sample to ensure that all samples were derived from the mid-blade. After harvesting, leaf samples were stored in Ziploc bags filled with water overnight at 4 °C, to ensure full hydration of epidermal cells and to capture an accurate representation of bulliform cell patterning under hydrated conditions.

Following hydration, leaf samples were pressed onto slides with Loctite® Super Glue Liquid Professional to generate leaf epidermal glue-impressions. Leaf glue-impressions were air-dried for at least 10 min, and removed from the leaf surfaces. Leaf epidermal glue-impressions were stored on slides at room temperature for future imaging. For each glue impression, three RGB images sampling different areas of the impression were taken with a Zeiss® Z1/ApoTome stereo-microscope in bright field using a 1X objective lens.

4.3.5 Neural networks in the quantification of phenotypic data

Convolutional neural networks (CNNs) were employed to quantify bulliform cell patterning traits in leaf epidermal glue-impression images. Each image was first resized to a 968 x 1292 grayscale image using Python module skimage 0.14.2 and cropped to the shape of 960 x 960 with Python module numpy 1.16.3. Each image was further split into four 480 x 480 images for faster computation. A training and validation set of 120 randomly sampled images and a test set of 20 randomly sampled images were created by manually annotating the pixels that are bulliform cells with Python module OpenCV 3.3.0 and skimage 0.14.2. Five U-nets were trained on 120 training images in Python with modules Keras 2.2.4 and TensorFlow 1.10.0. Ten percent of the 120 training images were used as the validation set for the determination of the optimal learning rate of 5×10^{-5}. Binary cross entropy was used as the loss function for the training, validation and test set. The output of five U-nets was aggregated as the finalized output segmentation map. After segmentation, every four 480 x 480 images were put back to their original 960 x 960 images to quantify the bulliform patterning phenotypes.

Each segmentation map is a two-dimensional array with binary elements. The two bulliform cell patterning phenotypes: bulliform cell column number and width, were quantified as below. In cases where there were more than three continuous pixels classified as bulliform cells, one column of bulliform cells was counted. The ratio between the total number of pixels annotated as bulliform cells and the number of bulliform cell columns is the average bulliform cell width of the image. To acquire model accuracies in regards to the bulliform cell patterning traits, a separate set of 30 images were manually annotated, and model accuracies were derived when comparing the CNN-generated segmentation map and the manual annotation.

4.3.6 Statistical data analysis

To screen the phenotypic data for significant outliers, mixed linear models were fitted as follows: (1) each single location; and (2) both locations. The model terms included grand mean and check as fixed effects and environment, genotype, genotype-by-environment (G×E) interaction (only for models ii), incomplete block within environment, and column within environment as random effects. The Studentized deleted residuals (Kutner *et al.*, 2005) generated from these mixed linear models were assessed and significant ($\alpha = 0.05$) outliers removed. For each outlier screened phenotype, an iterative mixed linear model fitting procedure was conducted for each of the two full models in ASReml-R version 3.0 (Butler *et al.*, 2009). All random terms that were not significant at $\alpha = 0.05$ in a likelihood ratio test were removed from the model, allowing a final best-fit model to be obtained for each phenotype. These final models were used to generate a best linear unbiased predictor (BLUP) for each line.

Variance component estimates from the fitted mixed linear models were fitted to allow for the estimation of variance components used to estimate heritability on a line-mean basis (Holland *et al.*, 2003; Hung *et al.*, 2012) for each phenotype within and across environments. Standard errors of the heritability estimates were calculated with the delta method (Lynch and Walsh, 1998; Holland *et al.*, 2003).

4.3.7 DNA extraction, genotyping and SNP identification

For each of the 468 inbred lines in the WiDiv panel, total genomic DNA was extracted from a bulk of young leaves from a single plant. The leaf tissue samples were lyophilized and ground using a GenoGrinder (Spex SamplePrep, Metuchen, NJ, USA), followed by the isolation of genomic DNA using the DNeasy 96 Plant Kit (Qiagen Inc., Valencia, CA, USA). DNA samples were sent for genotyping-by-sequencing (GBS) (Elshire *et al.*, 2011) at the Cornell Biotechnology Resource Center (Cornell University, Ithaca, NY, USA) with restriction enzyme *Ape*KI. GBS libraries were constructed and multiplexed 192-fold for sequencing on an Illumina NextSeq 500 instrument.

Genotypes at 955,690 high-confidence single-nucleotide polymorphism (SNP) loci were called with B73 RefGen_v2 coordinates as described (Baseggio *et al.*, 2019). The raw SNP genotype calls were filtered to discard singleton and doubleton SNPs (a minor allele observed in a single line), and only biallelic SNPs with call rates greater than 40% and minimum inbreeding coefficient of 0.8 were retained. Missing SNP genotypes were partially imputed using FILLIN (Swarts *et al.*, 2014) with a set of maize haplotype donor files with a 4 kb window size (AllZeaGBSv2.7impV5_AnonDonors4k.tar.gz, available at panzea.org). Physical

coordinates of the SNP loci were uplifted to B73 RefGen_AGPv4. To uplift physical coordinates of the SNP loci to B73 RefGen_AGPv4, a 101 bp flanking sequence for each SNP (+/- 50 bp from a SNP) was aligned to B73 RefGen_AGPv4 using Vmatch (Kurtz, 2003) to obtain the uplifted SNP coordinates. SNPs with flanking sequences that could not be uniquely and perfectly aligned to the reference genome were removed from the dataset. The final complete set contained 258,690 SNP markers.

4.3.8 Genome-wide association study

Identified SNPs with minimum minor allele counts of 40 (4.28% minor allele frequency), minimum call rates of 60%, maximum heterozygosity of 10%, and a minimal inbreeding coefficient of 0.8 were retained, resulting in 258,308 high-quality GBS SNP markers. For each bulliform cell patterning trait, a mixed linear model was used with R package GAPIT 3.0 enabling Population Parameters Previously Determined (P3D) to conduct the GWAS (Zhang *et al.*, 2010; Lipka *et al.*, 2012). A subset of 41,259 SNPs remaining after linkage disequilibrium (LD) pruning ($r^2 \leq 0.2$) of the complete marker data set in PLINK version 1.09_beta5 (Purcell *et al.*, 2007) were used to calculate the genomic relationship (kinship) matrix. The kinship matrix was calculated with the VanRaden method included in the GAPIT package with no compression used when conducting GWAS. After the removal of low quality images and outliers, 461 inbred lines were used in the GWAS analysis for each environment and across both environments. To control for the multiple testing problem. The false-discovery rate (FDR) was calculated as described in the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Significant association between the trait BLUPs and SNPs were tested and reported at the 5% FDR level.

4.3.9 Linkage disequilibrium analysis

Linkage disequilibrium (LD) was estimated with squared allele frequency correlations (r^2) as described (Lewontin, 1988). For each top (i.e., most significant) SNP at a locus, r^2 between all the other SNPs on the same chromosome and the top SNP were calculated, and genes that reside in a window spanned by SNPs in stronger than 0.5 LD with the top SNP were investigated as putative candidate genes.

4.3.10 Data availability

The raw GBS sequencing data were deposited at NCBI SRA with accession number SRP160407 and in BioProject under accession PRJNA489924. The raw RNAseq data were deposited at NCBI SRT with SRA accession numbers PRJNA545465 and PRJNA400334. Leaf epidermal glue-impression images can be found at https://de.cyverse.org/dl/d/8CA8D72B-24AF-4887-8899-14460021887A/resized.zip.

4.4 Results and Discussion

4.4.1 Bulliform cell ontogeny

The strap-like maize leaf is composed of the proximal sheath and the distal blade, which are separated by the ligule/auricle blade-sheath boundary (Figure 4.1). The sheath surrounds the stem and inserts at the node, whereas the blade extends away from the stem and is the major photosynthetic portion of the leaf. Bulliform cells are found only on the adaxial leaf blade, forming clusters that are 4-5 cells wide and arranged in linear columns that extend the length of the blade, in parallel to the lateral veins (Figure 1). Macrohairs are specialized hairs that develop in the center of the bulliform cell rows (Figure 4.1A).

The ontogeny of bulliform cells was investigated in order to generate an RNA sequencing (RNAseq) library from the site of bulliform cell initiation, to be used as a crosscheck of our GWAS candidate genes for bulliform cell patterning. At 33 days after planting, the maize B73 adult leaf number 8 is still elongating from a meristematic region near the base of the leaf blade, just distal to the ligule as shown in Figure 1B. Epidermal impressions near the proximal end of the leaf blade, approximately 2 mm distal to the ligule of maize leaf eight, show no morphological evidence of bulliform cell patterning (Figure 4.1B). Approximately 15 mm from the ligule, morphological differences in epidermal cells are observed (Figure 4.1C), although bulliform cells are not yet distinguishable. Thirty mm beyond the ligule however, cell types such as prickle hairs and bulliform cells are identified by their distinctive morphologies (Figure 4.1D). Thus, by proximally tracking bulliform cell rows that are visible at 30 mm from the ligule down to 15 mm from the ligule and lower, it is possible to identify immature bulliform cell rows before they develop their distinctive morphology. These analyses of epidermal cell development suggest that the bulliform cell ontogenic zone of the expanding leaf 8, where developmental patterning of the bulliform cells begins, is located approximately 2 mm above the ligule (Figure 4.1B).

RNAseq was performed on leaf tissue harvested from the bulliform cell ontogenic zone (Figure 4.1B). A differential gene expression analysis comparing the transcriptomes of the bulliform cell ontogenic zone and that of a distal blade interval harvested from 15mm -35mm above the ligule of leaf 8 was conducted. Using an FDR of < 0.10, 15,081 out of 18,264 total transcripts were differentially expressed in the bulliform cell ontogenic zone as compared to more the distal, differentiated leaf

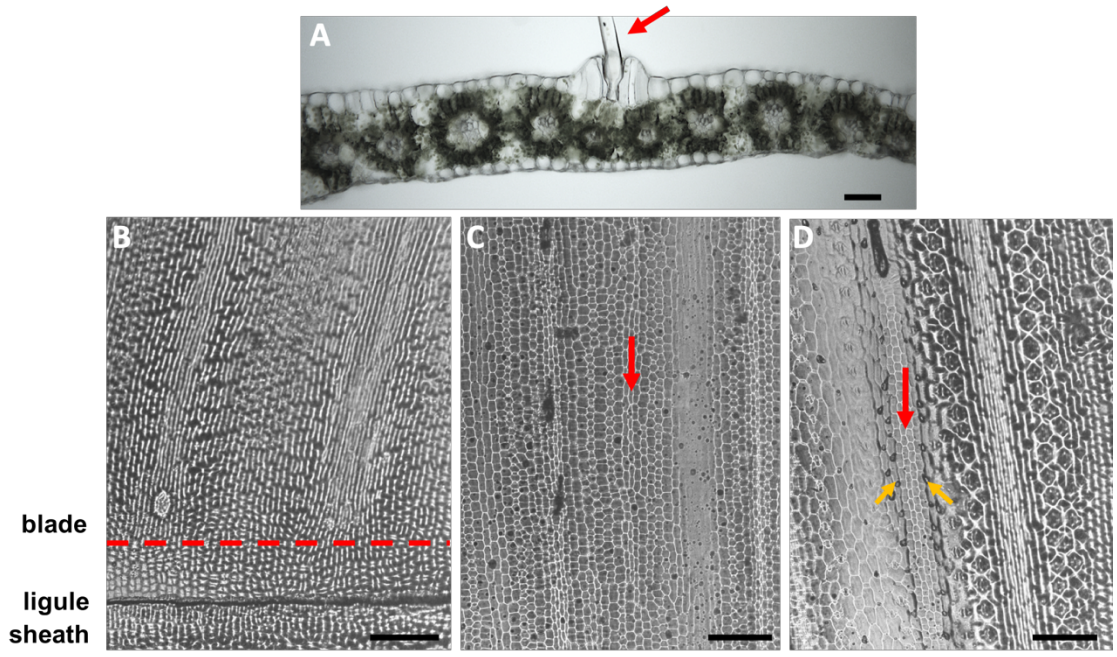

Figure 4.1. Ontogeny of maize bulliform cell development. (A) A mature bulliform cell cluster containing four morphologically distinct cells and a macrohair on the adaxial surface of an adult leaf. (B) No differences cellular morphology is detectable in a 2 mm region of blade immediately distal to the ligule (bounded by the dashed red line) of the emerging adult leaf 8. (C) 15 mm region distal to the ligule, showing differences in cell morphology in files of cell columns, but no distinguishing bulliform cell characteristics. (D) 30 mm region distal to the ligule. The red arrow in (D) marks the same bulliform cell column denoted by the red arrow in (C), which indicates the cell column in (C) is an early stage of bulliform cell ontogeny. Orange arrows denote prickle hairs flanking bulliform cells. Scale bar 200 μm in (A), 2 mm in (B-D).

tissues. These data suggest that bulliform cell patterning is regulated by a complex transcriptomic network. Importantly, this tissue-specific dataset provides a unique resource toward the selection of candidate genes contributing to bulliform cell patterning.

4.4.2 Phenotype variability and phenotyping accuracy

To survey the genetic diversity in maize bulliform cell patterning, leaf epidermal glue-impressions were obtained from the WiDiv panel, comprising 461 maize inbred lines grown in Maricopa, AZ, and San Diego, CA. Five glue-impressions per inbred line at each environment were sampled and three microscopic images were taken per glue-impression, for a total of 15,195 images. As shown in Figure 4.2, inbred lines comprising the WiDiv panel exhibit extreme variation in both bulliform column number and cell width (Table 4.1). To enable faster computation, each image was then subdivided into four segments. The resulting 60,780 sub-images were input to CNNs (U-nets) for computational identification (segmentation) of bulliform cells from the leaf epidermal glue-impressions. An output segmentation map, i.e. a binary grayscale image, was generated after the U-net segmented the raw images (Figure 4.3). The U-net model displayed an accuracy of 96.46% for bulliform column number, and 89.33% for bulliform column width.

Both bulliform cell patterning traits were highly heritable, indicating that these bulliform cell patterning traits have a strong genetic underpinning and are amenable to GWAS. Specifically, heritabilities on a line-mean basis for column number and width were 0.76 and 0.71 respectively across both environments, with plot-level heritability within each environment varying from 0.70 and 0.86 (Table 4.1).

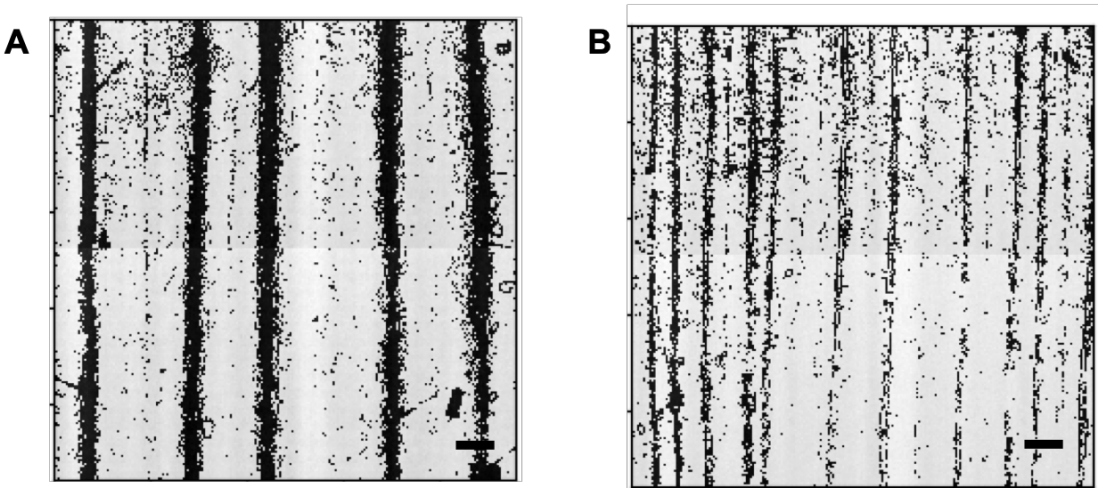

Figure 4.2. Grayscale images of leaf epidermal glue-impressions from two maize inbred lines showing extreme bulliform cell patterning phenotypes. (A) Inbred line MS153 shows 5 bulliform cell columns in this image, with an average width of 187.05 µm. (B) A374 has 11 bulliform cell columns with an average width of 63.57 µm. Scale bar 500 µm.

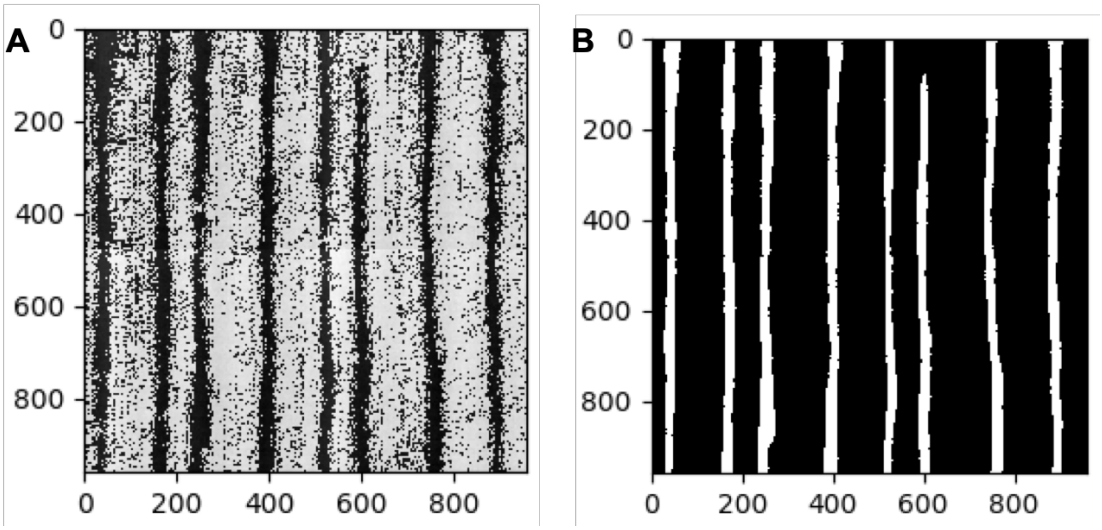

Figure 4.3. Segmentation output of U-nets from inbred line B79. (A) The raw image without annotation. (B) The segmentation map of the U-net output of the raw image in (A). In (B) white columns are bulliform cell columns; all other cells in the epidermal background are black. Each axis labels the pixels.

Table 4.1 Phenotypic diversity and heritability of bulliform patterning traits assessed

Traits	Number of Lines	BLUPs in environments combined			BLUPs in Maricopa, AZ			BLUPs in San Diego, CA		
		Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Column Number	461	9.35	0.7	7.15 - 11.79	9.67	0.84	7.42 - 12.36	8.91	0.72	6.60 - 11.13
Column Width	461	103.51	10.21	80.33 - 138.28	103.52	12.18	70.27 - 146.61	101.2	13.67	70.64 - 148.77

Traits	Heritabilities		
	Environments combined	Maricopa, AZ	San Diego, CA
Column Number	0.76 ± 0.024	0.86 ± 0.030	0.70 ± 0.066
Column Width	0.71 ± 0.029	0.81 ± 0.044	0.81 ± 0.041

4.4.3 GWAS of bulliform cell patterning traits

The genetic architecture of bulliform cell patterning traits was investigated with the WiDiv panel. GWAS results individually from Maricopa, AZ, San Diego, CA and combined results from both environments are summarized in Figure 4.4. A single SNP (give raw P value, give bp position) located on chromosome 4 is associate with bulliform column number at the 5% FDR level in the Maricopa environment. Although this same SNP is also the top SNP (i.e., most significant) associated with bulliform column number across environments, it is not significant at the 5% FDR level. In addition, this locus is not among the top SNPs for bulliform column number in San Diego, CA.

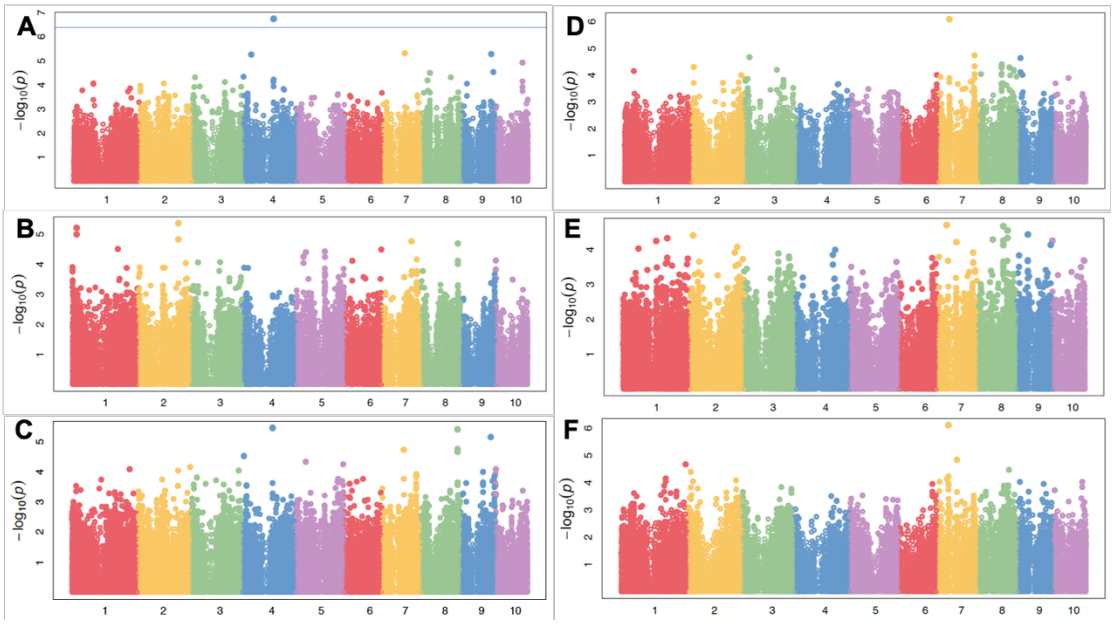

Figure 4.4. GWAS Manhattan plots for bulliform cell patterning traits. (A) Bulliform column number in Maricopa, AZ. The blue line indicates 0.05 FDR. (B) Bulliform column number in San Diego, CA. (C) Bulliform column number in both environments combined (Maricopa, AZ and San Diego, CA). (D) Bulliform column width in Maricopa, AZ. (E) Bulliform column width in San Diego, CA. (F) Bulliform column width in both environments combined (Maricopa, AZ and San Diego, CA).

To search for candidate genes regulating bulliform column number, we investigated LD of the peak SNP with nearby SNPs on chromosome 4; nine genes were identified within an 863.0 kb window having an r^2 greater than 0.5 with the peak SNP. However, just one of these candidate loci (*Zm00001d051057*) is transcriptionally upregulated in the bulliform ontogenic zone (Table 4.2), the predicted site of bulliform cell patterning. *Zm00001d051057* is predicted to encode a HISTONE-LYSINE N-METHYLTRANSFERASE and is homologous to the Arabidopsis gene *ASH1-RELATED3 (ASHR3*). *ASHR3* encodes a SET-domain protein conferring histone H3 lysine-36 methylation, with implicated functions during regulation of stem cell division in the root apical meristem (Kumpf *et al*., 2014). We speculate that in maize, this *HISTONE-LYSINE N-METHYLTRANSFERASE* homolog may regulate cell division in bulliform column initial cells.

In the same 863.0 kb region near the bulliform column number SNP, there are five genes downregulated in the bulliform ontogenic zone (Table 4.2), including *Zm00001d051065*, with a *CAP-BINDING PROTEIN 20* homolog that are implicated in the epidermal patterning in Arabidopsis (Jager *et al*., 2011), and a putative cell cycle gene homolog (*Zm00001d051061*) (Gelsthorpe *et al*., 1997). These comprise additional candidate genes regulating bulliform cell patterning.

Our GWAS identified a single top locus (located at 50129023 bp on chromosome 7, with raw p-values of $8.11 \times 10^{-7}, 7.77 \times 10^{-7}, 2.09 \times 10^{-4}$ in AZ, CA, and both environments combined, respectively) for bulliform cell column width (not significant at 5% FDR in any field location). The most significant SNP in Maricopa, AZ, and in both environments combined, this SNP is also highly ranked in San Diego, CA.

Among the 16 genes found to reside in a 1.93 Mb region spanned by SNPs having an r^2 greater than 0.5 with this top SNP, are transcriptionally upregulated in the bulliform ontogenic zone when compared to the bulliform maturation zone (Table 4.2). Notably, maize gene *Zm00001d019696* is predicted to encode a CYCLIN10 homolog, implicated to function during regulation of cell division. The Arabidopsis homolog CYCD3;2 mediates response to cytokinin, and regulates cell number in lateral organs (Dewitte *et al.*, 2007). Other candidate genes for bulliform cell width include a second predicted cyclin (*CYCLIN-LI-1*), as well as *Zm00001d019677* and *Zm00001d019688*. *Zm00001d019677* is predicted to encode a maize homolog of the Arabidopsis F-box protein VIER F-BOX PROTEIN1, whereas *Zm00001d019688* is homologous to the Arabidopsis gene *DEFECTIVE IN MERISTEM SILENCING 5* (*DMS5*) that functions in RNA-directed DNA methylation (Lopez Sanchez *et al*., 2016; Choudury *et al*., 2019). Intriguingly, the maize *ASHR3*-like gene, implicated above in our GWAS of bulliform row number, also functions in DNA methylation (Kumpf *et al*., 2014). These data suggest that bulliform cell patterning may be epigenetically regulated.

Table 4.2. Bulliform cell patterning candidate genes identified

Trait	Candidate Gene	Upregulated in bulliform ontogenic zone	Maize Gene Name	Arabidopsis Gene Name
Column Number	Zm00001d051057	TRUE	*ASHR3*	*ASH1-RELATED 3*
Column Width	Zm00001d019696	TRUE	*CYCLIN10*	*CYCD3;2*
Column Width	Zm00001d019677	TRUE	*NA*	*VIER F-BOX PROTEIN 1*
Column Width	Zm00001d019688	TRUE	*NA*	*DEFECTIVE IN MERISTEM SILENCING 5*
Column Width	Zm00001d019681	TRUE	*CYCLIN-L1-1*	*ARGININE-RICH CYCLIN 1*
Column Number	Zm00001d051055	FALSE	*CYTOCHROME C OXIDASE POLYPEPTIDE*	*CYTOCHROME OXIDASE*
Column Number	Zm00001d051062	FALSE	*GRPE PROTEIN HOMOLOG*	*CHLOROPLAST GRPE 1*
Column Number	Zm00001d051063	FALSE	*PHOSPHATIDYL-N-METHYLETHANOLAMINE N-METHYLTRANSFERASE*	*ARABIDOPSIS PHOSPHOLIPID N-METHYLTRANSFERASE*
Column Number	Zm00001d051061	FALSE	*ENHANCER OF RUDIMENTARY HOMOLOG*	*ARABIDOPSIS THALIANA ENHANCER OF RUDIMENTARY HOMOLOGUE*
Column Number	Zm00001d051065	FALSE	*NUCLEAR CAP-BINDING PROTEIN SUBUNIT 2*	*CAP-BINDING PROTEIN 20*

Despite the high heritability of the bulliform cell patterning traits described in this study, few statistically-associated GWAS hits are identified. Several factors may contribute to this phenomenon. For example, bulliform cell patterning may be regulated by many genes with relatively small effects, which our mapping population may have insufficient statistical power to detect. In addition, these phenotypes could also be controlled by rare alleles in the population that were filtered out before conducting the GWAS. Lastly, extremely diverse environments may have dramatic effects on bulliform cell patterning phenotypes, thus why the strongest associations were mainly identified in the Maricopa environment. Plants grown in Maricopa, AZ are predicted to undergo extreme water conservation responses, as compared to the same inbred lines raised in the mild climate of San Diego, CA. Additional environmental replicates may help dissect the genotype-by-environment effects of this complex trait.

This study combines developmental analyses and stage-specific transcriptomics with the high-throughput microscopic phenotyping power enabled by machine learning, together with quantitative genetics and genomics, to investigate the genetic architecture of bulliform cell patterning. Although a microscopic phenotype, bulliform cell patterning is an important agronomic trait with implications in macroscopic phenotypes such as plant architecture and drought resistance. We identify five candidate genes in the regulation of bulliform column number and width. Future reverse genetic analyses, and transcriptomic studies of bulliform cell patterning mutants, can further investigate the roles of these candidate genes in this important yet understudied trait.

4.5 Acknowledgements

The work was funded by NSF IOS award #1444507. We thank all members of the cuticle project for discussion and input.

4.6 REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Baseggio, M., Murray, M., Magallanes-Lundback, M., Kaczmar, N., Chamness, J., Buckler, E.S., Smith, M.E., DellaPenna, D., Tracy, W.F., and Gore, M.A. (2019). Genome-Wide Association and Genomic Prediction Models of Tocochromanols in Fresh Sweet Corn Kernels. Plant Genome **12**.

Becraft, P.W., Li, K., Dey, N., and Asuncion-Crabb, Y. (2002). The maize dek1 gene functions in embryonic pattern formation and cell fate specification. Development **129,** 5217-5225.

Benjamini, Y., and Hochberg, Y.J.J.o.t.R.s.s.s.B. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing **57,** 289-300.

Bennetzen, J.L., and Hake, S.C. (2008). Handbook of maize: its biology. (Springer Science & Business Media).

Butler, D., Cullis, B.R., Gilmour, A., Gogel, B.J.T.S.o.Q., Department of Primary Industries, and Fisheries, B. (2009). ASReml-R reference manual.

Chen, Q., Xie, Q., Gao, J., Wang, W., Sun, B., Liu, B., Zhu, H., Peng, H., Zhao, H., Liu, C., Wang, J., Zhang, J., Zhang, G., and Zhang, Z. (2015). Characterization of Rolled and Erect Leaf 1 in regulating leave morphology in rice. J Exp Bot **66,** 6047-6058.

Choudury, S.G., Shahid, S., Cuerda-Gil, D., Panda, K., Cullen, A., Ashraf, Q., Sigman, M.J., McCue, A.D., and Slotkin, R.K. (2019). The RNA Export Factor

ALY1 Enables Genome-Wide RNA-Directed DNA Methylation. Plant Cell **31,** 759-774.

Dai, M., Zhao, Y., Ma, Q., Hu, Y., Hedden, P., Zhang, Q., and Zhou, D.-X.J.P.p. (2007). The rice YABBY1 gene is involved in the feedback regulation of gibberellin metabolism **144,** 121-133.

Dewitte, W., Scofield, S., Alcasabas, A.A., Maughan, S.C., Menges, M., Braun, N., Collins, C., Nieuwland, J., Prinsen, E., Sundaresan, V., and Murray, J.A. (2007). Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. Proc Natl Acad Sci U S A **104,** 14537-14542.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6,** e19379.

Fang, L., Zhao, F., Cong, Y., Sang, X., Du, Q., Wang, D., Li, Y., Ling, Y., Yang, Z., and He, G. (2012). Rolling-leaf14 is a 2OG-Fe (II) oxygenase family protein that modulates rice leaf rolling by affecting secondary cell wall formation in leaves. Plant Biotechnol J **10,** 524-532.

Fujino, K., Matsuda, Y., Ozawa, K., Nishimura, T., Koshiba, T., Fraaije, M.W., Sekiguchi, H.J.M.G., and Genomics. (2008). NARROW LEAF 7 controls leaf shape mediated by auxin in rice **279,** 499-507.

Gelsthorpe, M., Pulumati, M., McCallum, C., Dang-Vu, K., and Tsubota, S.I. (1997). The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. Gene **186,** 189-195.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

Hibara, K., Obara, M., Hayashida, E., Abe, M., Ishimaru, T., Satoh, H., Itoh, J., and Nagato, Y. (2009). The ADAXIALIZED LEAF1 gene functions in leaf and embryonic pattern formation in rice. Dev Biol **334,** 345-354.

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2014). Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26,** 121-135.

Holland, J.B., Nyquist, W.E., and Cervantes-Martínez, C.T.J.P.b.r. (2003). Estimating and interpreting heritability for plant breeding: an update **22,** 9-112.

Hsiao, T.C., O'Toole, J.C., Yambao, E.B., and Turner, N.C.J.P.P. (1984). Influence of osmotic adjustment on leaf rolling and tissue death in rice (Oryza sativa L.) **75,** 338-341.

Hu, J., Zhu, L., Zeng, D., Gao, Z., Guo, L., Fang, Y., Zhang, G., Dong, G., Yan, M., and Liu, J.J.P.m.b. (2010). Identification and characterization of NARROW ANDROLLED LEAF 1, a novel gene regulating leaf morphology and plant architecture in rice **73,** 283-292.

Hung, H., Browne, C., Guill, K., Coles, N., Eller, M., Garcia, A., Lepak, N., Melia-Hancock, S., Oropeza-Rosas, M., and Salvo, S.J.H. (2012). The relationship

between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population **108,** 490.

Itoh, J., Hibara, K., Sato, Y., and Nagato, Y. (2008). Developmental role and auxin responsiveness of Class III homeodomain leucine zipper gene family members in rice. Plant Physiol **147,** 1960-1975.

Jager, K., Fabian, A., Tompa, G., Deak, C., Hohn, M., Olmedilla, A., Barnabas, B., and Papp, I. (2011). New phenotypes of the drought-tolerant cbp20 Arabidopsis thaliana mutant have changed epidermal morphology. Plant Biol (Stuttg) **13,** 78-84.

Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H.Y., Zhao, H., Deng, X.W., and Nelson, T. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet **41,** 258-263.

Kadioglu, A., and Terzi, R.J.T.B.R. (2007). A dehydration avoidance mechanism: leaf rolling **73,** 290-302.

Kenrick, P., and Crane, P.R.J.N. (1997). The origin and early evolution of plants on land **389,** 33.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods **12,** 357-360.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097-1105.

Kumpf, R., Thorstensen, T., Rahman, M.A., Heyman, J., Nenseth, H.Z., Lammens, T., Herrmann, U., Swarup, R., Veiseth, S.V., Emberland, G., Bennett, M.J., De Veylder, L., and Aalen, R.B. (2014). The ASH1-RELATED3 SET-domain protein controls cell division competence of the meristem and the quiescent center of the Arabidopsis primary root. Plant Physiol **166,** 632-643.

Kurtz, S.J.R.T.C.P. (2003). The Vmatch large scale sequence analysis software **412,** 297.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). Applied linear statistical models. (McGraw-Hill Irwin Boston).

LeCun, Y., Bengio, Y.J.T.h.o.b.t., and networks, n. (1995). Convolutional networks for images, speech, and time series **3361,** 1995.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.J.P.o.t.I. (1998). Gradient-based learning applied to document recognition **86,** 2278-2324.

Lewontin, R.C. (1988). On measures of gametic disequilibrium. Genetics **120,** 849-852.

Li, L., Shi, Z.Y., Li, L., Shen, G.Z., Wang, X.Q., An, L.S., and Zhang, J.L. (2010). Overexpression of ACL1 (abaxially curled leaf 1) increased Bulliform cells and induced Abaxial curling of leaf blades in rice. Mol Plant **3,** 807-817.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics **28,** 2397-2399.

Lopez Sanchez, A., Stassen, J.H., Furci, L., Smith, L.M., and Ton, J. (2016). The role of DNA (de)methylation in immune responsiveness of Arabidopsis. Plant J **88,** 361-374.

Lynch, M., and Walsh, B. (1998). Genetics and analysis of quantitative traits. (Sinauer Sunderland, MA).

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res **40,** 4288-4297.

Ort, D.R., and Long, S.P. (2014). Botany. Limits on yields in the Corn Belt. Science **344,** 484-485.

Price, A.H., Young, E., and Tomos, A.J.T.N.P. (1997). Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (Oryza sativa) **137,** 83-91.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J.J.T.A.j.o.h.g. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses **81,** 559-575.

Raven, J.A., and Edwards, D. (2004). Physiological evolution of lower embryophytes: adaptations to the terrestrial environment. In The evolution of plant physiology (Elsevier), pp. 17-41.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (Springer), pp. 234-241.

Simonyan, K., and Zisserman, A.J.a.p.a. (2014). Very deep convolutional networks for large-scale image recognition.

Swarts, K., Li, H., Romero Navarro, J.A., An, D., Romay, M.C., Hearne, S., Acharya, C., Glaubitz, J.C., Mitchell, S., and Elshire, R.J.J.T.P.G. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants **7**.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9.

Xiang, J.-J., Zhang, G.-H., Qian, Q., and Xue, H.-W.J.P.p. (2012). Semi-rolled leaf1 encodes a putative glycosylphosphatidylinositol-anchored protein and modulates rice leaf rolling by regulating the formation of bulliform cells **159,** 1488-1500.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and

Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet **38,** 203-208.

Zeiler, M.D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (Springer), pp. 818-833.

Zhang, G.H., Xu, Q., Zhu, X.D., Qian, Q., and Xue, H.W. (2009). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. Plant Cell **21,** 719-735.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat Genet **42,** 355-360.

Zou, L.-p., Sun, X.-h., Zhang, Z.-g., Liu, P., Wu, J.-x., Tian, C.-j., Qiu, J.-l., and Lu, T.-g.J.P.p. (2011). Leaf rolling controlled by the homeodomain leucine zipper class IV gene Roc5 in rice **156,** 1589-1602.

Chapter 5: Conclusion

In this dissertation, the epidermal patterning of the maize leaf was investigated with regard to its role in the control of water loss. Different scales of data were exploited, from the epidermal-specific transcriptome to the wide diversity of epidermal glue-impression images from nearly 500 maize inbred lines. This thesis research is focused on three topics related to drought tolerance and water conservation; (1) cuticle development, (2) cuticular evaporation rate (CE), and (3) bulliform cell patterning, as summarized below.

Cuticles seal the above-ground shoots of plants, where their hydrophobic properties prevent water loss (Kenrick and Crane, 1997a; Raven and Edwards, 2004; Raven *et al.*, 2005). Many genes/proteins involved in cuticle biosynthesis are previously identified, as are some of the complex mechanisms that regulate cuticle biogenesis (Suh *et al.*, 2005a; Kunst and Samuels, 2009b; Yeats and Rose, 2013a). Previous studies demonstrated that light performs a vital role in upregulation of cuticle biosynthesis (Tobin and Silverthorne, 1985; Fankhauser and Chory, 1997; Mittmann *et al.*, 2004). Here we established the first mechanistic link between PHYTOCHROME-mediated light signaling and cuticle development, and suggest a model whereby light influences the lipid production and transport in the leaf epidermis. Further experiments in maize, moss, and Arabidopsis provide support for the hypothesis that LIPID-TRANSFER PROTEINs comprise one evolutionary novelty that enabled cuticle formation in land plants. In addition, this study generated tissue-specific and stage-specific transcriptomic data that will bolster many avenues of plant

research, such as the study of bulliform cell patterning described in Chapter 4 of this dissertation.

Although a critical innovation in embryophyte evolution, cuticles are not an absolute seal against water loss from plant leaves. For example, water loss that occurs after stomatal closure, or cuticular evaporation (CE), comprises the majority of water loss for plants under drought conditions (Kelliher *et al.*, 1995; Daszkowska-Golec and Szarejko, 2013). In the third chapter of this dissertation, the natural diversity of CE rate was measured and exploited for genome- and transcriptome-wide association studies (GWAS and TWAS), to identify candidate genes regulating CE rate in maize. Subsequently, GWAS and TWAS were combined and synergized to boost the statistical power of candidate gene identification. Functional information in the gene expression data, generated by Gene Ontology analyses, provided further support for the role of vesicular trafficking as an important regulator for CE rate in maize. Cuticles components are synthesized in the endoplasmic reticulum (ER), and subsequently exported from the plasma membrane to the epidermal cell surface. Nevertheless, the transport route for cuticle lipids from the ER and the plasma membrane has not been explicitly demonstrated. Candidate genes identified in our GWAS and TWAS analyses provide support for the involvement of vesicular trafficking in the transport of cuticle lipid components from the ER to the apoplast. These gene candidates comprise plausible reverse genetic targets for investigations of cuticle function, toward crop improvement.

Aside from the cuticle, land plants have evolved additional strategies to prevent water loss, such as the development of bulliform cell rows on the epidermis of grass leaves

(Bennetzen and Hake, 2008; Robinson *et al.*, 2015). Bulliform cells are enlarged, parenchymatous cells that shrink in size during water deficit, which causes leaf rolling to reduce water evaporation (Hsiao *et al.*, 1984; Price *et al.*, 1997; Dai *et al.*, 2007; Kadioglu and Terzi, 2007; Hu *et al.*, 2010). We exploited the natural diversity of bulliform cell patterning, sampled as bulliform row number and row width, to identify candidate genes regulating bulliform cell patterning. An ensemble of convolutional neural networks was employed to annotate the segmentation of bulliform cell rows, compared to all other adaxial epidermal cell types. GWAS analyses performed were to identify candidate genes associated with these AI-derived bulliform cell patterning traits, which implicated genes involved in cell division and DNA methylation.

Although comprehensive, this dissertation leaves many interesting questions for further exploration. For example, does the *LTP* homolog identified in Chapter 2 transport wax components of specific chain lengths? Biochemical analysis such as chromatography can test the affinity of different lipid components for this LTP, whereas genetic analysis of null mutations for this maize gene may also address this question. In order to test the correlation of PHY-mediated light signaling and cuticle evolution, light/dark treatments similar to the one described in Chapter 2 can be performed on moss and algae, followed by differential gene expression analysis. To investigate the environmental effects on bulliform cell patterning, more trails within the same location could provide the statistical power to calculate G×E, and phenotypic data collected in more environments may provide a more accurate estimate of the bulliform patterning phenotypes sampled in Chapter 4.

In summary, this dissertation straddles the interface of multiple research disciplines, including developmental biology, quantitative genetics and genomics, biostatistics, and machine learning, and makes important contributions to our understanding of cuticle biogenesis and function, and bulliform cell patterning. We expect that the multiple, unique datasets generated by this research dissertation will provide a stimulus to future studies of water-loss conservation in crop plants.

5.1 REFERENCES

Bennetzen, J.L., and Hake, S.C. (2008). Handbook of maize: its biology. (Springer Science & Business Media).

Dai, M., Zhao, Y., Ma, Q., Hu, Y., Hedden, P., Zhang, Q., and Zhou, D.-X.J.P.p. (2007). The rice YABBY1 gene is involved in the feedback regulation of gibberellin metabolism **144,** 121-133.

Daszkowska-Golec, A., and Szarejko, I. (2013). Open or close the gate - stomata action under the control of phytohormones in drought stress conditions. Front Plant Sci **4,** 138.

Fankhauser, C., and Chory, J. (1997). Light control of plant development. Annu Rev Cell Dev Biol **13,** 203-229.

Hsiao, T.C., O'Toole, J.C., Yambao, E.B., and Turner, N.C.J.P.P. (1984). Influence of osmotic adjustment on leaf rolling and tissue death in rice (Oryza sativa L.) **75,** 338-341.

Hu, J., Zhu, L., Zeng, D., Gao, Z., Guo, L., Fang, Y., Zhang, G., Dong, G., Yan, M., and Liu, J.J.P.m.b. (2010). Identification and characterization of NARROW ANDROLLED LEAF 1, a novel gene regulating leaf morphology and plant architecture in rice **73,** 283-292.

Kadioglu, A., and Terzi, R.J.T.B.R. (2007). A dehydration avoidance mechanism: leaf rolling **73,** 290-302.

Kelliher, F., Leuning, R., Raupach, M., Schulze, E.-D.J.A., and Meteorology, F. (1995). Maximum conductances for evaporation from global vegetation types **73,** 1-16.

Kenrick, P., and Crane, P.R.J.N. (1997). The origin and early evolution of plants on land **389,** 33.

Kunst, L., and Samuels, L.J.C.o.i.p.b. (2009). Plant cuticles shine: advances in wax biosynthesis and export **12,** 721-727.

Mittmann, F., Brucker, G., Zeidler, M., Repp, A., Abts, T., Hartmann, E., and Hughes, J. (2004). Targeted knockout in Physcomitrella reveals direct actions of phytochrome in the cytoplasm. Proc Natl Acad Sci U S A **101,** 13939-13944.

Price, A.H., Young, E., and Tomos, A.J.T.N.P. (1997). Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (Oryza sativa) **137,** 83-91.

Raven, J.A., and Edwards, D. (2004). Physiological evolution of lower embryophytes: adaptations to the terrestrial environment. In The evolution of plant physiology (Elsevier), pp. 17-41.

Raven, P.H., Evert, R.F., and Eichhorn, S.E. (2005). Biology of plants. (Macmillan).

Robinson, D.O., Roeder, A.H.J.C.O.i.G., and Development. (2015). Themes and variations in cell type patterning in the plant epidermis **32,** 55-65.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiol **139,** 1649-1665.

Tobin, E.M., and Silverthorne, J.J.A.R.o.P.P. (1985). Light regulation of gene expression in higher plants **36,** 569-593.

Yeats, T.H., and Rose, J.K.J.P.p. (2013). The formation and function of plant cuticles **163,** 5-20.

Chapter 6 Appendix: Transcriptomic analyses of genes regulating SAM size in maize

6.1 Abstract

The morphology of the maize shoot apical meristem (SAM) correlates with important agronomic traits such as flowering time and stem width. However, comparative transcriptomic studies of different SAM morphotypes are lacking. In this study, tissue-specific transcriptomes were analyzed from SAMs of distinct sizes classes, to identify mRNAs implicated in the regulation of SAM morphological variation in maize. MADS-box gene transcripts constitute more than one third of the differentially expressed mRNAs. This study provides a transcriptomic dataset that is amenable to genetic analyses of SAM morphometric variation.

6.2 Introduction

The maize shoot apical meristem (SAM) is a pool of stem cells that gives rise to all above-ground tissues. Heritable morphological differences in SAM size and shape have been observed among maize inbred lines. SAM morphological variation correlates with important agronomic traits such as flowering time, internode width, plant height, and leaf number (Leiboff *et al.*, 2015). In this study, tissue-specific, comparative transcriptomics are utilized to examine the quantitative transcriptional differences that correlate with morphometric variation in the maize SAM. Candidates genes regulating SAM size are identified, many of which are implicated in the regulation of flowering time. Reverse genetic analyses of the genes identified in this

study may also contribute to a better understanding of the genetic mechanisms regulating adult plant architecture in maize.

6.3 Materials and Methods

6.3.1 Plant material and growth conditions

Inbred lines comprising large SAMs (Co255, F42, ND246), medium SAMs (B73, Mo17, W22) and small SAMs (B64, CML322, B57) (Leiboff *et al.*, 2015) were obtained from the Maize Genetics Cooperation Stock Center, and grown in a Percival A100 growth chamber for two weeks. Growth conditions were 25 °C day, 20 °C night, 60% relative humidity, and 10-hour day length.

6.3.2 Laser-microdissection RNA sequencing (LM-RNAseq)

Shoot apices of two-week old seedlings were harvested, fixed, and paraplast-embedded for use in laser microsdissection as described (Takacs *et al.*, 2012). Following SAM microdissection, RNA was extracted using the PicoPure™ RNA isolation kit, and linearly-amplified using the TargetAmp™ 2-Round aRNA Amplification Kit 2.0. The NEBNext Ultra™ RNA Library Prep Kit for Illumina was used for construction of RNA sequencing libraries, which were sequenced using Illumina HiSEQ 2500 instrument. RNAseq reads were trimmed with Trimmomatic 0.35 (Bolger *et al.*, 2014), aligned with Tophat2 (Kim *et al.*, 2013), and counted with HTSeq (Anders *et al.*, 2015).

6.3.3 Differential gene expression analysis

RNAseq reads obtained from the three inbred lines representing each, distinct SAM size class (i.e. the large SAMs Co255, F42, ND246, the medium SAMs B73, Mo17, W22, and the small SAMs B64, CML322, B57) were pooled for analyses of differential gene expression. Group-wise comparisons were made with R package edgeR. Genes with false discovery rate of 0.05 were declared as significantly differentially expressed.

6.4 Results and Discussion

6.4.1 Differential gene expression analysis

Inbred lines sampled were grouped into three categories: large (Co255, F42, ND246), medium (B73, Mo17, W22), and small (B64, B57, CML322). For each of these nine inbred lines, eight to ten shoot apical meristems (SAMs) were laser-microdissected (Figure 6.1) as described (Brooks III *et al.*, 2009). Eight to ten SAMs were pooled as one RNA sequencing (RNAseq) sample, and three biological replicates of each inbred line were sequenced. A total of 22,701 reads were identified with more than one count per million reads. Three inbred lines were pooled to comprise each SAM size category, and group-wise comparisons were made between large and medium SAMs, large and small SAMs, and medium and small SAMs.

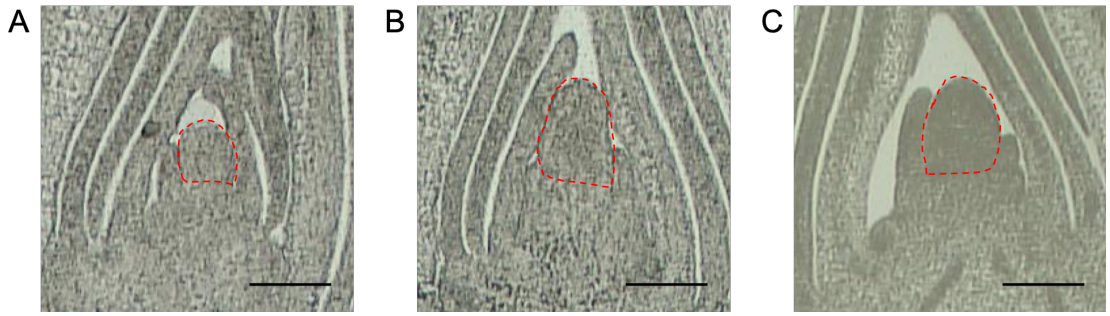

Figure 6.1. Images of SAMs for laser-microdissection. Images of a small SAM from B57 (A), a medium SAM from B73 (B), and a large SAM from Co255 are shown. The red-dotted lines indicate the sample areas that were laser-microdissected. Scale bars: 100 μm

The numbers of differentially-expressed genes are summarized in Figure 6.2. Among the three group-wise comparison, large vs. small SAMs have the most differentially-expressed (DE) genes (2530), while large and medium SAMs contain the fewest (588). Surprisingly, just 22 genes were commonly DE, eleven (Table 6.1) of which exhibit the same direction of fold change when comparing large to medium and medium to small SAMs. For example, gene *GRMZM2G032339* is upregulated in large vs. medium SAMs, and also upregulated medium vs. small SAMs. Four of these eleven genes are predicted to encode transcription factors with homology to Arabidopsis MADS-box proteins implicated in the regulation of flowering time or floral meristem identity (Purugganan *et al.*, 1995; Becker *et al.*, 2003). Interestingly, previous studies showed that SAM size in maize is indeed negatively correlated with flowering time (Leiboff *et al.*, 2015). It is intriguing that these seedling SAMs accumulate mRNAs implicated to function during flowering, although at the time of harvest none of the two week-old SAMs in this study showed morphological evidence of the transition to inflorescence meristems.

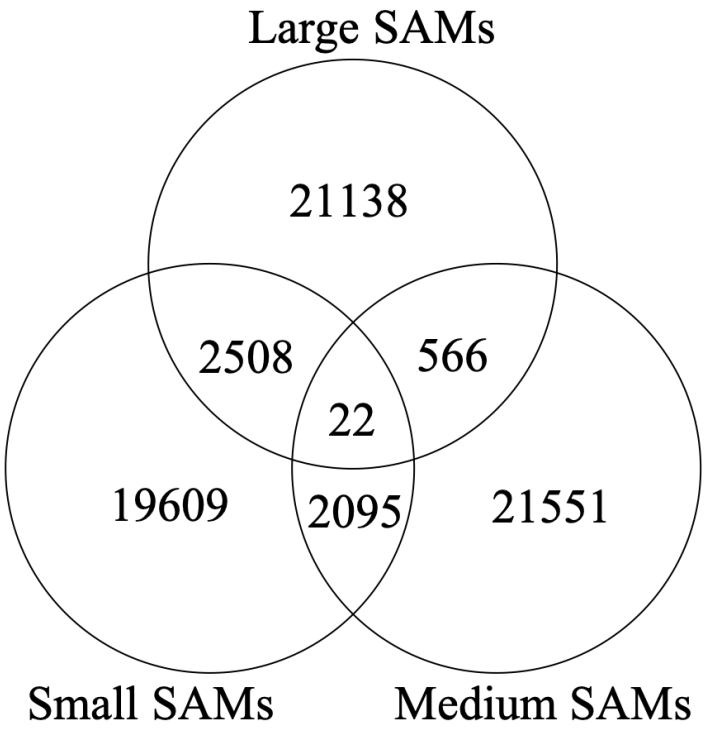

Figure 6.2. Venn diagram shows overlapping differentially expressed genes in three groups. Numbers in overlapping circles represent the number of differentially expressed genes. For example, 22 genes were identified differentially expressed in all three comparisons.

Table 6.1. Eleven overlapping differentially expressed genes that are upregulated in larger SAMs

Maize gene ID	Maize gene annotation	Arabidopsis homolog	Arabidopsis homolog name
GRMZM2G032339	*MADS-TRANSCRIPTION FACTOR 4*	*AT1G69120*	*AGAMOUS-LIKE 7*
GRMZM2G009987	*NA*	*AT5G58290*	*REGULATORY PARTICLE TRIPLE-A ATPASE 3*
GRMZM2G053757	*NA*		
GRMZM2G077131	*NA*	*AT4G20010*	*ORGANELLAR SINGLE-STRANDED DNA BINDING PROTEIN 2*
GRMZM2G126566	*MYB-TRANSCRIPTION FACTOR 159*		
GRMZM2G115812	*NA*	*AT1G76740*	
GRMZM2G165044	*NA*	*AT2G15530*	
GRMZM2G553379	*MADS-BOX 15*	*AT1G69120*	*AGAMOUS-LIKE 7*
GRMZM2G064382	*NA*	*AT1G10970*	*ZINC TRANSPORTER 4 PRECURSOR*
GRMZM2G072582	*MADS-BOX 3*	*AT1G69120*	*AGAMOUS-LIKE 7*
GRMZM2G148693	*ZEA APETALA HOMOLOG1*	*AT1G69120*	*AGAMOUS-LIKE 7*

6.5 REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Becker, A., Theißen, G.J.M.p., and evolution. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants **29,** 464-489.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30,** 2114-2120.

Brooks III, L., Strable, J., Zhang, X., Ohtsu, K., Zhou, R., Sarkar, A., Hargreaves, S., Elshire, R.J., Eudy, D., and Pawlowska, T.J.P.g. (2009). Microdissection of shoot meristem functional domains **5,** e1000476.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14,** R36.

Leiboff, S., Li, X., Hu, H.-C., Todt, N., Yang, J., Li, X., Yu, X., Muehlbauer, G.J., Timmermans, M.C., and Yu, J.J.N.c. (2015). Genetic control of morphometric diversity in the maize shoot apical meristem **6,** 8974.

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J., and Yanofsky, M.F.J.G. (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family **140,** 345-356.

Takacs, E.M., Li, J., Du, C., Ponnala, L., Janick-Buckner, D., Yu, J., Muehlbauer, G.J., Schnable, P.S., Timmermans, M.C., Sun, Q., Nettleton, D., and

Scanlon, M.J. (2012). Ontogeny of the maize shoot apical meristem. Plant Cell **24,** 3219-3234.

BIBLIOGRAPHY

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Ashton, D.J.A.J.o.B. (1975). The root and shoot development of Eucalyptus regnans F. Muell **23,** 867-887.

Barabasi, A.-L. (2003). Linked: How everything is connected to everything else and what it means.

Barabasi, A.L. (2013). Network science. Philos Trans A Math Phys Eng Sci **371,** 20120375.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet **5,** 101-113.

Baseggio, M., Murray, M., Magallanes-Lundback, M., Kaczmar, N., Chamness, J., Buckler, E.S., Smith, M.E., DellaPenna, D., Tracy, W.F., and Gore, M.A. (2019). Genome-Wide Association and Genomic Prediction Models of Tocochromanols in Fresh Sweet Corn Kernels. Plant Genome **12**.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., and Stein, W.E. (1998a). Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. Annual Review of Ecology and Systematics **29,** 263-292.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., Stein, W.E.J.A.R.o.E., and Systematics. (1998b). Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation **29,** 263-292.

Beaudoin, F., Wu, X., Li, F., Haslam, R.P., Markham, J.E., Zheng, H., Napier, J.A., and Kunst, L. (2009). Functional characterization of the Arabidopsis beta-ketoacyl-coenzyme A reductase candidates of the fatty acid elongase. Plant Physiol **150,** 1174-1191.

Becker, A., Theißen, G.J.M.p., and evolution. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants **29,** 464-489.

Becraft, P.W., Li, K., Dey, N., and Asuncion-Crabb, Y. (2002a). The maize dek1 gene functions in embryonic pattern formation and cell fate specification. Development **129,** 5217-5225.

Becraft, P.W., Li, K., Dey, N., and Asuncion-Crabb, Y.J.D. (2002b). The maize dek1 gene functions in embryonic pattern formation and cell fate specification **129,** 5217-5225.

Benjamini, Y., and Hochberg, Y.J.J.o.t.R.s.s.s.B. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing **57,** 289-300.

Bennetzen, J.L., and Hake, S.C. (2008). Handbook of maize: its biology. (Springer Science & Business Media).

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.D., Haslam, R.P., Napier, J.A., Lessire, R., and Joubes, J. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that Arabidopsis ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. Plant Cell **24,** 3106-3118.

Bernardo, R. (2002). Breeding for quantitative traits in plants. (Stemma press Woodbury).

Bird, D., Beisson, F., Brigham, A., Shin, J., Greer, S., Jetter, R., Kunst, L., Wu, X., Yephremov, A., and Samuels, L. (2007). Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J **52,** 485-498.

Block, H., Knight Jr, B., and Rosenblatt, F.J.R.o.M.P. (1962). Analysis of a four-layer series-coupled perceptron. II **34,** 135.

Blum, A., and Rivest, R.L. (1989). Training a 3-node neural network is NP-complete. In Advances in neural information processing systems, pp. 494-501.

Boke, N.H.J.B. (1980). Developmental morphology and anatomy in Cactaceae **30,** 605-610.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30,** 2114-2120.

Bourgault, R., Matschi, S., Vasquez, M., Qiao, P., Sonntag, A., Charlebois, C., Mohammadi, M., Scanlon, M.J., Smith, L.G., and Molina, I. (2019a). Changes in lipid composition and ultrastructure associated with functional maturation of the cuticle during adult maize leaf development**,** 625343.

Bourgault, R., Matschi, S., Vasquez, M., Qiao, P., Sonntag, A., Charlebois, C., Mohammadi, M., Scanlon, M.J., Smith, L.G., and Molina, I.J.b. (2019b). Changes in lipid composition and ultrastructure associated with functional maturation of the cuticle during adult maize leaf development**,** 625343.

Boyer, J.S.J.A.r.o.p.p. (1985). Water transport **36,** 473-516.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S.J.B. (2007). TASSEL: software for association mapping of complex traits in diverse samples **23,** 2633-2635.

Brooks III, L., Strable, J., Zhang, X., Ohtsu, K., Zhou, R., Sarkar, A., Hargreaves, S., Elshire, R.J., Eudy, D., and Pawlowska, T.J.P.g. (2009). Microdissection of shoot meristem functional domains **5,** e1000476.

Broun, P., Poindexter, P., Osborne, E., Jiang, C.-Z., and Riechmann, J.L.J.P.o.t.N.A.o.S. (2004). WIN1, a transcriptional activator of epidermal wax accumulation in Arabidopsis **101,** 4706-4711.

Browning, B.L., and Browning, S.R.J.T.A.J.o.H.G. (2016). Genotype imputation with millions of reference samples **98,** 116-126.

Browning, B.L., Zhou, Y., and Browning, S.R.J.T.A.J.o.H.G. (2018). A one-penny imputed genome from next-generation reference panels **103,** 338-348.

Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., Wang, B., Xu, D., Yang, B., and Xie, C.J.G. (2017). Construction of the third-generation Zea mays haplotype map **7,** gix134.

Butler, D., Cullis, B.R., Gilmour, A., Gogel, B.J.T.S.o.Q., Department of Primary Industries, and Fisheries, B. (2009). ASReml-R reference manual.

Chen, Q., Xie, Q., Gao, J., Wang, W., Sun, B., Liu, B., Zhu, H., Peng, H., Zhao, H., and Liu, C.J.J.o.e.b. (2015a). Characterization of Rolled and Erect Leaf 1 in regulating leave morphology in rice **66,** 6047-6058.

Chen, Q., Xie, Q., Gao, J., Wang, W., Sun, B., Liu, B., Zhu, H., Peng, H., Zhao, H., Liu, C., Wang, J., Zhang, J., Zhang, G., and Zhang, Z. (2015b). Characterization of Rolled and Erect Leaf 1 in regulating leave morphology in rice. J Exp Bot **66,** 6047-6058.

Chen, Y.A., and Scheller, R.H.J.N.r.M.c.b. (2001). SNARE-mediated membrane fusion **2,** 98.

Chory, J., and Susek, R.E.J.C.S.H.M.A. (1994). 22 Light Signal Transduction and the Control of Seedling Development **27,** 579-614.

Choudury, S.G., Shahid, S., Cuerda-Gil, D., Panda, K., Cullen, A., Ashraf, Q., Sigman, M.J., McCue, A.D., and Slotkin, R.K. (2019). The RNA Export Factor ALY1 Enables Genome-Wide RNA-Directed DNA Methylation. Plant Cell **31,** 759-774.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16,** 735-743.

Coll, E.P., Kandt, C., Bird, D.A., Samuels, A.L., and Tieleman, D.P.J.T.J.o.P.C.B. (2007). The distribution and conformation of very long-chain plant wax components in a lipid bilayer **111,** 8702-8704.

Coppen, J.J. (2003). Eucalyptus: the genus Eucalyptus. (CRC Press).

Cowan, I.J.J.o.A.E. (1965). Transport of water in the soil-plant-atmosphere system**,** 221-239.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133,** 462-469.

Dai, M., Zhao, Y., Ma, Q., Hu, Y., Hedden, P., Zhang, Q., and Zhou, D.-X.J.P.p. (2007). The rice YABBY1 gene is involved in the feedback regulation of gibberellin metabolism **144,** 121-133.

Daszkowska-Golec, A., and Szarejko, I. (2013). Open or close the gate - stomata action under the control of phytohormones in drought stress conditions. Front Plant Sci **4,** 138.

Debono, A., Yeats, T.H., Rose, J.K., Bird, D., Jetter, R., Kunst, L., and Samuels, L. (2009). Arabidopsis LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. Plant Cell **21,** 1230-1238.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (Ieee), pp. 248-255.

Dewey, M. (2016). metap: Meta-analysis of significance values. R package version 0.7.

Dewitte, W., Scofield, S., Alcasabas, A.A., Maughan, S.C., Menges, M., Braun, N., Collins, C., Nieuwland, J., Prinsen, E., Sundaresan, V., and Murray, J.A. (2007). Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. Proc Natl Acad Sci U S A **104,** 14537-14542.

Duanmu, D., Bachy, C., Sudek, S., Wong, C.H., Jimenez, V., Rockwell, N.C., Martin, S.S., Ngan, C.Y., Reistetter, E.N., van Baren, M.J., Price, D.C., Wei, C.L., Reyes-Prieto, A., Lagarias, J.C., and Worden, A.Z. (2014). Marine algae and

land plants share conserved phytochrome signaling systems. Proc Natl Acad Sci U S A **111,** 15827-15832.

Edqvist, J., Blomqvist, K., Nieuwland, J., and Salminen, T.A. (2018). Plant lipid transfer proteins: are we finally closing in on the roles of these enigmatic proteins? J Lipid Res **59,** 1374-1382.

Edstam, M.M., Viitanen, L., Salminen, T.A., and Edqvist, J. (2011). Evolutionary history of the non-specific lipid transfer proteins. Mol Plant **4,** 947-964.

Edwards, D.J.N.P. (1993). Cells and tissues in the vegetative sporophytes of early land plants **125,** 225-247.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6,** e19379.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., and Zisserman, A.J.I.j.o.c.v. (2010). The pascal visual object classes (voc) challenge **88,** 303-338.

Fang, L., Zhao, F., Cong, Y., Sang, X., Du, Q., Wang, D., Li, Y., Ling, Y., Yang, Z., and He, G. (2012a). Rolling-leaf14 is a 2OG-Fe (II) oxygenase family protein that modulates rice leaf rolling by affecting secondary cell wall formation in leaves. Plant Biotechnol J **10,** 524-532.

Fang, L., Zhao, F., Cong, Y., Sang, X., Du, Q., Wang, D., Li, Y., Ling, Y., Yang, Z., and He, G.J.P.b.j. (2012b). Rolling-leaf14 is a 2OG-Fe (II) oxygenase family protein that modulates rice leaf rolling by affecting secondary cell wall formation in leaves **10,** 524-532.

Fankhauser, C., and Chory, J. (1997). Light control of plant development. Annu Rev Cell Dev Biol **13,** 203-229.

Fich, E.A., Segerson, N.A., and Rose, J.K.J.A.R.o.P.B. (2016). The plant polyester cutin: biosynthesis, structure, and biological roles **67,** 207-233.

Finkina, E.I., Melnikova, D.N., Bogdanov, I.V., and Ovchinnikova, T.V. (2016). Lipid Transfer Proteins As Components of the Plant Innate Immune System: Structure, Functions, and Applications. Acta Naturae **8,** 47-61.

Fisher, R.A.J.E., Scotland. (1925). Statistical methods for research workers Oliver and Boyd **6**.

Fujino, K., Matsuda, Y., Ozawa, K., Nishimura, T., Koshiba, T., Fraaije, M.W., Sekiguchi, H.J.M.G., and Genomics. (2008). NARROW LEAF 7 controls leaf shape mediated by auxin in rice **279,** 499-507.

Gelsthorpe, M., Pulumati, M., McCallum, C., Dang-Vu, K., and Tsubota, S.I. (1997). The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. Gene **186,** 189-195.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S.J.P.o. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline **9,** e90346.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. (MIT press).

Gore, M.A., Chia, J.-M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills, G.S., and Ross-Ibarra, J.J.S. (2009). A first-generation haplotype map of maize **326,** 1115-1117.

Gurel, P.S., Hatch, A.L., and Higgs, H.N. (2014). Connecting the cytoskeleton to the endoplasmic reticulum and Golgi. Curr Biol **24,** R660-R672.

Gyorffy, B., Hatzis, C., Sanft, T., Hofstatter, E., Aktas, B., and Pusztai, L. (2015). Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Res **17,** 11.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (Los Alamos National Lab.(LANL), Los Alamos, NM (United States)).

Hansey, C.N., Johnson, J.M., Sekhon, R.S., Kaeppler, S.M., and Leon, N.d.J.C.S. (2011). Genetic diversity of a maize association population with restricted phenology **51,** 704-715.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In Neural networks for perception (Elsevier), pp. 65-93.

Hehnly, H., and Stamnes, M. (2007). Regulating cytoskeleton-based vesicle motility. FEBS Lett **581,** 2112-2118.

Hibara, K., Obara, M., Hayashida, E., Abe, M., Ishimaru, T., Satoh, H., Itoh, J., and Nagato, Y. (2009a). The ADAXIALIZED LEAF1 gene functions in leaf and embryonic pattern formation in rice. Dev Biol **334,** 345-354.

Hibara, K.-i., Obara, M., Hayashida, E., Abe, M., Ishimaru, T., Satoh, H., Itoh, J.-i., and Nagato, Y.J.D.b. (2009b). The ADAXIALIZED LEAF1 gene functions in leaf and embryonic pattern formation in rice **334,** 345-354.

Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F., Lindquist, E., Pedraza, M.A., Barry, K., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2014). Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26,** 121-135.

Holland, J.B., Nyquist, W.E., and Cervantes-Martínez, C.T.J.P.b.r. (2003). Estimating and interpreting heritability for plant breeding: an update **22,** 9-112.

Hooker, T.S., Millar, A.A., and Kunst, L. (2002). Significance of the expression of the CER6 condensing enzyme for cuticular wax production in Arabidopsis. Plant Physiol **129,** 1568-1580.

Hooker, T.S., Lam, P., Zheng, H., and Kunst, L.J.T.P.C. (2007). A core subunit of the RNA-processing/degrading exosome specifically influences cuticular wax biosynthesis in Arabidopsis **19,** 904-913.

Horvath, S. (2011). Weighted network analysis: applications in genomics and systems biology. (Springer Science & Business Media).

Hsiao, T.C., O'Toole, J.C., Yambao, E.B., and Turner, N.C.J.P.P. (1984). Influence of osmotic adjustment on leaf rolling and tissue death in rice (Oryza sativa L.) **75,** 338-341.

Hu, J., Zhu, L., Zeng, D., Gao, Z., Guo, L., Fang, Y., Zhang, G., Dong, G., Yan, M., and Liu, J.J.P.m.b. (2010). Identification and characterization of NARROW

ANDROLLED LEAF 1, a novel gene regulating leaf morphology and plant architecture in rice **73,** 283-292.

Hung, H., Browne, C., Guill, K., Coles, N., Eller, M., Garcia, A., Lepak, N., Melia-Hancock, S., Oropeza-Rosas, M., and Salvo, S.J.H. (2012). The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population **108,** 490.

Itoh, J., Hibara, K., Sato, Y., and Nagato, Y. (2008a). Developmental role and auxin responsiveness of Class III homeodomain leucine zipper gene family members in rice. Plant Physiol **147,** 1960-1975.

Itoh, J.-I., Hibara, K.-I., Sato, Y., and Nagato, Y.J.P.P. (2008b). Developmental role and auxin responsiveness of Class III homeodomain leucine zipper gene family members in rice **147,** 1960-1975.

Jager, K., Fabian, A., Tompa, G., Deak, C., Hohn, M., Olmedilla, A., Barnabas, B., and Papp, I. (2011). New phenotypes of the drought-tolerant cbp20 Arabidopsis thaliana mutant have changed epidermal morphology. Plant Biol (Stuttg) **13,** 78-84.

Jetter, R., and Riederer, M.J.P.p. (2016a). Localization of the transpiration barrier in the epi-and intracuticular waxes of eight plant species: water transport resistances are associated with fatty acyl rather than alicyclic components **170,** 921-934.

Jetter, R., and Riederer, M. (2016b). Localization of the Transpiration Barrier in the Epi- and Intracuticular Waxes of Eight Plant Species: Water Transport Resistances Are Associated with Fatty Acyl Rather Than Alicyclic Components. Plant Physiol **170,** 921-934.

Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H.Y., Zhao, H., Deng, X.W., and Nelson, T. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet **41,** 258-263.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A., and O'Brien, S.J. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics **11,** 724.

Johnston, R., Wang, M., Sun, Q., Sylvester, A.W., Hake, S., and Scanlon, M.J. (2014). Transcriptomic analyses indicate that maize ligule development recapitulates gene expression patterns that occur during lateral organ initiation. Plant Cell **26,** 4718-4732.

Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R.J.P.m.b. (2008). The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling **67,** 547.

Kadioglu, A., and Terzi, R.J.T.B.R. (2007). A dehydration avoidance mechanism: leaf rolling **73,** 290-302.

Kawaguchi, K. (2016). Deep learning without poor local minima. In Advances in neural information processing systems, pp. 586-594.

Kelliher, F., Leuning, R., Raupach, M., Schulze, E.-D.J.A., and Meteorology, F. (1995). Maximum conductances for evaporation from global vegetation types **73,** 1-16.

Kenrick, P., and Crane, P.R.J.N. (1997a). The origin and early evolution of plants on land **389,** 33.

Kenrick, P., and Crane, P.R. (1997b). The origin and early evolution of plants on land. Nature **389,** 33.

Kerstiens, G. (2006a). Water transport in plant cuticles: an update. J Exp Bot **57,** 2493-2499.

Kerstiens, G.J.J.o.E.B. (2006b). Water transport in plant cuticles: an update **57,** 2493-2499.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods **12,** 357-360.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14,** R36.

Kim, H., Go, Y.S., and Suh, M.C. (2018). DEWAX2 Transcription Factor Negatively Regulates Cuticular Wax Biosynthesis in Arabidopsis Leaves. Plant Cell Physiol **59,** 966-977.

Kim, H., Lee, S.B., Kim, H.J., Min, M.K., Hwang, I., and Suh, M.C. (2012). Characterization of glycosylphosphatidylinositol-anchored lipid transfer protein 2 (LTPG2) and overlapping function between LTPG/LTPG1 and LTPG2 in cuticular wax export or accumulation in Arabidopsis thaliana. Plant Cell Physiol **53,** 1391-1403.

Kolattukudy, P.E. (2001a). Polyesters in higher plants. In Biopolyesters (Springer), pp. 1-49.

Kolattukudy, P.E. (2001b). Polyesters in higher plants. Adv Biochem Eng Biotechnol **71,** 1-49.

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. Plant Methods **9,** 29.

Kosma, D.K., Rice, A., and Pollard, M. (2015). Analysis of aliphatic waxes associated with root periderm or exodermis from eleven plant species. Phytochemistry **117,** 351-362.

Kremling, K., Diepenbrock, C., Gore, M., Buckler, E., and Bandillo, N.J.b. (2018a). Transcriptome-wide association supplements genome-wide association in Zea mays**,** 363242.

Kremling, K.A.G., Chen, S.Y., Su, M.H., Lepak, N.K., Romay, M.C., Swarts, K.L., Lu, F., Lorant, A., Bradbury, P.J., and Buckler, E.S. (2018b). Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature **555,** 520-523.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097-1105.

Kumpf, R., Thorstensen, T., Rahman, M.A., Heyman, J., Nenseth, H.Z., Lammens, T., Herrmann, U., Swarup, R., Veiseth, S.V., Emberland, G., Bennett, M.J., De Veylder, L., and Aalen, R.B. (2014). The ASH1-RELATED3 SET-domain protein controls cell division competence of the meristem and the quiescent center of the Arabidopsis primary root. Plant Physiol **166,** 632-643.

Kunst, L., and Samuels, L. (2009a). Plant cuticles shine: advances in wax biosynthesis and export. Curr Opin Plant Biol **12,** 721-727.

Kunst, L., and Samuels, L.J.C.o.i.p.b. (2009b). Plant cuticles shine: advances in wax biosynthesis and export **12,** 721-727.

Kurdyukov, S., Faust, A., Trenkamp, S., Bar, S., Franke, R., Efremova, N., Tietjen, K., Schreiber, L., Saedler, H., and Yephremov, A. (2006). Genetic and biochemical evidence for involvement of HOTHEAD in the biosynthesis of long-chain alpha-,omega-dicarboxylic fatty acids and formation of extracellular matrix. Planta **224,** 315-329.

Kurtz, S.J.R.T.C.P. (2003). The Vmatch large scale sequence analysis software **412,** 297.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). Applied linear statistical models. (McGraw-Hill Irwin Boston).

Lam, P., Zhao, L., McFarlane, H.E., Aiga, M., Lam, V., Hooker, T.S., and Kunst, L.J.P.p. (2012). RDR1 and SGS3, components of RNA-mediated gene silencing, are required for the regulation of cuticular wax biosynthesis in developing inflorescence stems of Arabidopsis **159,** 1385-1395.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9,** 559.

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw **46**.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics **24,** 719-720.

LeCun, Y., Bengio, Y.J.T.h.o.b.t., and networks, n. (1995). Convolutional networks for images, speech, and time series **3361,** 1995.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.J.P.o.t.I. (1998). Gradient-based learning applied to document recognition **86,** 2278-2324.

Lee, S.B., Go, Y.S., Bae, H.J., Park, J.H., Cho, S.H., Cho, H.J., Lee, D.S., Park, O.K., Hwang, I., and Suh, M.C. (2009). Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules, and enhanced susceptibility to infection by the fungal pathogen Alternaria brassicicola. Plant Physiol **150,** 42-54.

Leiboff, S., Li, X., Hu, H.-C., Todt, N., Yang, J., Li, X., Yu, X., Muehlbauer, G.J., Timmermans, M.C., and Yu, J.J.N.c. (2015). Genetic control of morphometric diversity in the maize shoot apical meristem **6,** 8974.

Lewontin, R.C. (1988). On measures of gametic disequilibrium. Genetics **120,** 849-852.

Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., and Kunst, L.J.P.p. (2008). Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis **148,** 97-107.

Li, F.W., Melkonian, M., Rothfels, C.J., Villarreal, J.C., Stevenson, D.W., Graham, S.W., Wong, G.K., Pryer, K.M., and Mathews, S. (2015). Phytochrome diversity in green plants and the origin of canonical plant phytochromes. Nat Commun **6,** 7852.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27,** 2987-2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25,** 2078-2079.

Li, L., Shi, Z.Y., Li, L., Shen, G.Z., Wang, X.Q., An, L.S., and Zhang, J.L. (2010a). Overexpression of ACL1 (abaxially curled leaf 1) increased Bulliform cells and induced Abaxial curling of leaf blades in rice. Mol Plant **3,** 807-817.

Li, L., Shi, Z.-Y., Li, L., Shen, G.-Z., Wang, X.-Q., An, L.-S., and Zhang, J.-L.J.M.p. (2010b). Overexpression of ACL1 (abaxially curled leaf 1) increased bulliform cells and induced abaxial curling of leaf blades in rice **3,** 807-817.

Li, Q., Hu, J., Ding, J., and Zheng, G. (2014). Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. Biostatistics **15,** 284-295.

Lin, H.Y., Liu, Q., Li, X., Yang, J., Liu, S., Huang, Y., Scanlon, M.J., Nettleton, D., and Schnable, P.S. (2017). Substantial contribution of genetic variation in the

expression of transcription factors to phenotypic variation revealed by eRD-GWAS. Genome Biol **18,** 192.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics **28,** 2397-2399.

Lopez Sanchez, A., Stassen, J.H., Furci, L., Smith, L.M., and Ton, J. (2016). The role of DNA (de)methylation in immune responsiveness of Arabidopsis. Plant J **88,** 361-374.

Lü, S., Zhao, H., Des Marais, D.L., Parsons, E.P., Wen, X., Xu, X., Bangarusamy, D.K., Wang, G., Rowland, O., and Juenger, T.J.P.p. (2012). Arabidopsis ECERIFERUM9 involvement in cuticle formation and maintenance of plant water status **159,** 930-944.

Lynch, M., and Walsh, B. (1998). Genetics and analysis of quantitative traits. (Sinauer Sunderland, MA).

Mansbach, C.M., and Siddiqi, S.A.J.A.r.o.p. (2010). The biogenesis of chylomicrons **72,** 315-333.

Martin, M.J.E.j. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads **17,** 10-12.

Mauseth, J.D.J.B.o.t.T.B.C. (1995). Collapsible water-storage cells in cacti**,** 145-151.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res **40,** 4288-4297.

McFarlane, H.E., Shin, J.J., Bird, D.A., and Samuels, A.L. (2010). Arabidopsis ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. Plant Cell **22,** 3066-3075.

McFarlane, H.E., Watanabe, Y., Yang, W., Huang, Y., Ohlrogge, J., and Samuels, A.L.J.P.p. (2014). Golgi-and trans-Golgi network-mediated vesicle trafficking is required for wax secretion from epidermal cells **164,** 1250-1260.

Mittmann, F., Brucker, G., Zeidler, M., Repp, A., Abts, T., Hartmann, E., and Hughes, J. (2004). Targeted knockout in Physcomitrella reveals direct actions of phytochrome in the cytoplasm. Proc Natl Acad Sci U S A **101,** 13939-13944.

Molz, F.J.J.W.r.r. (1981). Models of water transport in the soil-plant system: A review **17,** 1245-1260.

Monneveux, P., and Belhassen, E. (1996). The diversity of drought adaptation in the wide. In Drought Tolerance in Higher Plants: Genetical, Physiological and Molecular Biological Analysis (Springer), pp. 7-14.

Nawrath, C. (2002a). The biopolymers cutin and suberin. Arabidopsis Book **1,** e0021.

Nawrath, C.J.T.A.b.A.S.o.P.B. (2002b). The biopolymers cutin and suberin **1**.

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. Nature **456,** 720-723.

Ogburn, R.M., and Edwards, E.J. (2010). The ecological water-use strategies of succulent plants. In Advances in botanical research (Elsevier), pp. 179-225.

Ort, D.R., and Long, S.P. (2014). Botany. Limits on yields in the Corn Belt. Science **344,** 484-485.

Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet **18,** 117-127.

Penfold, C.A., and Wild, D.L. (2011). How to infer gene networks from expression profiles, revisited. Interface Focus **1,** 857-870.

Pighin, J.A., Zheng, H., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L., and Samuels, A.L. (2004). Plant cuticular lipid export requires an ABC transporter. Science **306,** 702-704.

Price, A.H., Young, E., and Tomos, A.J.T.N.P. (1997). Quantitative trait loci associated with stomatal conductance, leaf rolling and heading date mapped in upland rice (Oryza sativa) **137,** 83-91.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J.J.T.A.j.o.h.g. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses **81,** 559-575.

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J., and Yanofsky, M.F.J.G. (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family **140,** 345-356.

Raven, J.A., and Edwards, D. (2004). Physiological evolution of lower embryophytes: adaptations to the terrestrial environment. In The evolution of plant physiology (Elsevier), pp. 17-41.

Raven, P.H., Evert, R.F., and Eichhorn, S.E. (2005). Biology of plants. (Macmillan).

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S.J.P.o.t.N.A.o.S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome **98,** 11479-11484.

Reyes, F.C., Buono, R.A., Roschzttardtz, H., Di Rubbo, S., Yeun, L.H., Russinova, E., and Otegui, M.S. (2014). A novel endosomal sorting complex required for transport (ESCRT) component in Arabidopsis thaliana controls cell expansion and development. J Biol Chem **289,** 4980-4988.

Rincent, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., Rodriguez, V.M., Moreno-Gonzalez, J., Melchinger, A., and Bauer, E.J.G. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (Zea mays L.) **192,** 715-728.

Robinson, D.O., Roeder, A.H.J.C.O.i.G., and Development. (2015). Themes and variations in cell type patterning in the plant epidermis **32,** 55-65.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (Springer), pp. 234-241.

Rowland, O., Lee, R., Franke, R., Schreiber, L., and Kunst, L. (2007). The CER3 wax biosynthetic gene from Arabidopsis thaliana is allelic to WAX2/YRE/FLP1. FEBS Lett **581,** 3538-3544.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., and Bernstein, M.J.I.j.o.c.v. (2015). Imagenet large scale visual recognition challenge **115,** 211-252.

Salminen, T.A., Blomqvist, K., and Edqvist, J. (2016). Lipid transfer proteins: classification, nomenclature, structure, and function. Planta **244,** 971-997.

Salminen, T.A., Eklund, D.M., Joly, V., Blomqvist, K., Matton, D.P., and Edqvist, J. (2018). Deciphering the Evolution and Development of the Cuticle by Studying Lipid Transfer Proteins in Mosses and Liverworts. Plants (Basel) **7**.

Samuels, L., Kunst, L., and Jetter, R. (2008). Sealing plant surfaces: cuticular wax formation by epidermal cells. Annu Rev Plant Biol **59,** 683-707.

Schaefer, R.J., Michno, J.M., Jeffers, J., Hoekenga, O., Dilkes, B., Baxter, I., and Myers, C.L. (2018). Integrating Coexpression Networks with GWAS to Prioritize Causal Genes in Maize. Plant Cell **30,** 2922-2942.

Sheehan, M.J., Kennedy, L.M., Costich, D.E., and Brutnell, T.P. (2007). Subfunctionalization of PhyB1 and PhyB2 in the control of seedling and mature plant traits in maize. Plant J **49,** 338-353.

Simonyan, K., and Zisserman, A.J.a.p.a. (2014). Very deep convolutional networks for large-scale image recognition.

Sorigue, D., Legeret, B., Cuine, S., Morales, P., Mirabella, B., Guedeney, G., Li-Beisson, Y., Jetter, R., Peltier, G., and Beisson, F. (2016). Microalgae Synthesize Hydrocarbons from Long-Chain Fatty Acids via a Light-Dependent Pathway. Plant Physiol **171,** 2393-2405.

Sorigue, D., Legeret, B., Cuine, S., Blangy, S., Moulin, S., Billon, E., Richaud, P., Brugiere, S., Coute, Y., Nurizzo, D., Muller, P., Brettel, K., Pignol, D., Arnoux, P., Li-Beisson, Y., Peltier, G., and Beisson, F. (2017). An algal photoenzyme converts fatty acids to hydrocarbons. Science **357,** 903-907.

Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput Biol **6,** e1000770.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc **7,** 500-507.

Suh, M.C., and Go, Y.S. (2014). DEWAX-mediated transcriptional repression of cuticular wax biosynthesis in Arabidopsis thaliana. Plant Signal Behav **9,** e29463.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005a). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiol **139,** 1649-1665.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005b). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiology **139,** 1649-1665.

Swarts, K., Li, H., Romero Navarro, J.A., An, D., Romay, M.C., Hearne, S., Acharya, C., Glaubitz, J.C., Mitchell, S., and Elshire, R.J.J.T.P.G. (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants **7**.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9.

Takacs, E.M., Li, J., Du, C., Ponnala, L., Janick-Buckner, D., Yu, J., Muehlbauer, G.J., Schnable, P.S., Timmermans, M.C., Sun, Q., Nettleton, D., and Scanlon, M.J. (2012). Ontogeny of the maize shoot apical meristem. Plant Cell **24,** 3219-3234.

Tobin, E.M., and Silverthorne, J.J.A.R.o.P.P. (1985). Light regulation of gene expression in higher plants **36,** 569-593.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend, S.H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature **415,** 530-536.

Vandepoele, K., Quimbaya, M., Casneuf, T., De Veylder, L., and Van de Peer, Y. (2009). Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. Plant Physiol **150,** 535-546.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am J Hum Genet **90,** 7-24.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet **101,** 5-22.

Whitelam, G.C., and Halliday, K.J. (2007). Light and plant development. (Wiley Online Library).

Whitewoods, C.D., Cammarata, J., Nemec Venza, Z., Sang, S., Crook, A.D., Aoyama, T., Wang, X.Y., Waller, M., Kamisugi, Y., Cuming, A.C., Szovenyi, P., Nimchuk, Z.L., Roeder, A.H.K., Scanlon, M.J., and Harrison, C.J. (2018). CLAVATA Was a Genetic Novelty for the Morphological Innovation of 3D Growth in Land Plants. Curr Biol **28,** 2365-2376 e2365.

Wilson, A.G. (2014). Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes (University of Cambridge).

Wu, R., Li, S., He, S., Wassmann, F., Yu, C., Qin, G., Schreiber, L., Qu, L.J., and Gu, H. (2011). CFL1, a WW domain protein, regulates cuticle development by modulating the function of HDG1, a class IV homeodomain transcription factor, in rice and Arabidopsis. Plant Cell **23,** 3392-3411.

Xiang, J.-J., Zhang, G.-H., Qian, Q., and Xue, H.-W.J.P.p. (2012). Semi-rolled leaf1 encodes a putative glycosylphosphatidylinositol-anchored protein and modulates rice leaf rolling by regulating the formation of bulliform cells **159,** 1488-1500.

Yeats, T.H., and Rose, J.K.J.P.p. (2013a). The formation and function of plant cuticles **163,** 5-20.

Yeats, T.H., and Rose, J.K. (2013b). The formation and function of plant cuticles. Plant Physiol **163,** 5-20.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet **38,** 203-208.

Zeiler, M.D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (Springer), pp. 818-833.

Zhang, G.-H., Xu, Q., Zhu, X.-D., Qian, Q., and Xue, H.-W.J.T.P.C. (2009a). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development **21,** 719-735.

Zhang, G.H., Xu, Q., Zhu, X.D., Qian, Q., and Xue, H.W. (2009b). SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. Plant Cell **21,** 719-735.

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat Genet **42,** 355-360.

Zhao, S.Q., Hu, J., Guo, L.B., Qian, Q., and Xue, H.W. (2010). Rice leaf inclination2, a VIN3-like protein, regulates leaf angle through modulating cell division of the collar. Cell Res **20,** 935-947.

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L., and Ye, Z.H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell **20,** 2763-2782.

Zou, L.-p., Sun, X.-h., Zhang, Z.-g., Liu, P., Wu, J.-x., Tian, C.-j., Qiu, J.-l., and Lu, T.-g.J.P.p. (2011). Leaf rolling controlled by the homeodomain leucine zipper class IV gene Roc5 in rice **156,** 1589-1602.